DOCUMENT RESUME

ED 041 105 VT 010 177

TITLE	Psychological Assessment of Patrolman Qualifications in Relation to Field Performance; The Identification of Predictors for Overall Performance of Patrolmen and the Relation between Predictors and Specific Patterns of Exceptional and Marginal Performance.
INSTITUTION	Chicago Police Dept., Ill.
SPONS AGENCY	Department of Justice, Washington, D.C. Office of Law Enforcement Assistance.
PUB DATE	5 Nov 68
NOTE	250p.
AVAILABLE FROM	Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402 (J1.2:P27/2, $2.00)
EDRS PRICE	EDRS Price MF-$1.00 HC Not Available from EDRS.
DESCRIPTORS	*Behavioral Science Research, *Behavior Patterns, *Law Enforcement, *Performance Criteria, *Personnel Selection

ABSTRACT

Behavioral scientists have been slow in the development of effective procedures and in empirical validation of psychological testing instruments for selecting law enforcement personnel. This study involved 253 supervisory sergeants and field lieutenants and 2,327 uniformed patrolmen of the Chicago Police Department (CPD). Three major performance measures were tenure, a paired-comparison performance rating which stressed field performance, and the CPD semiannual performance rating which stressed administrative as well as field performance. An occupational test battery, which stressed behavioral requirements of the patrolman's job, provided motivational, intellectual, and behavioral measures. The test battery was administered twice, with a 5-month interval, to allow for validation analysis. Patrolmen who scored high on the tests were rated high on performance. The degree of this relationship increased when the subgroup of white patrolmen was treated separately and increased even more significantly when the Negro subgroup was treated separately. It was concluded that ideal attributes of patrolmen are all related to stability (1) in parental and personal family life, (2) in self-confidence and emotional behavior, (3) in maintaining cooperative attitudes, and (4) in developing a realistic orientation toward life. (CH)

Law Enforcement Assistance Administration / U.S. Department of Justice / Washington, D.C.

Psychological Assessment of Patrolman Qualifications In Relation to Field Performance / LEAA Project #046

ED041105

VT010177

ED041105

November 5, 1968

PSYCHOLOGICAL ASSESSMENT OF PATROLMAN QUALIFICATIONS IN RELATION TO FIELD PERFORMANCE

The Identification of Predictors for Overall Performance of Patrolmen and the Relation between Predictors and Specific Patterns of Exceptional and Marginal Performance

Report Submitted to

Office of Law Enforcement Assistance
United States Department of Justice

Melany E. Baehr, Ph. D., Project Director
John E. Furcon, Field Director
Ernest C. Froemel, Statistical Coordinator
The Industrial Relations Center
The University of Chicago

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

This Project Supported by Grant #046
Awarded by the Attorney General
Under The Law Enforcement Assistance Act of 1965
to
The Chicago Police Department
Chicago, Illinois

Persons undertaking such projects under Goverment sponsorship are encouraged to express freely their professional judgments, findings, and conclusions. Therefore, points of view or opinions stated in this document do not necessarily represent the official position or policy of the U.S. Department of Justice.

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402 - Price $2.00

ACKNOWLEDGEMENTS

The authors would like to express their sincere thanks to Colonel John C. Bucher, Director of Personnel, Chicago Police Department, and Richard A. Heffernan, Assistant Director of Personnel, for their unstinting support and encouragement during the conduct of this project. The study was initiated under the auspices of then-Superintendent Orlando W. Wilson, and the complete cooperation of the Chicago Police Department has continued to the present with the support of Superintendent James B. Conlisk, Jr. The professional support and counsel provided by Frances M. Burns, David R. Saunders, Ph.D., Burton A. Grossman, Ph.D., and Charles R. Potkay, Ph.D., is deeply appreciated. Ronald G. Franczak, Arnold Oppenheim, and Anthony F. Jurkus were particularly instrumental in the data collection and analysis phases of the study. The project could not have been completed without the strong secretarial support provided by Donna C. Tanzer, Michelle A. Gregorich, Linda C. Thomas,Elaine Marzal, and Mary Ann Chalupa of the Industrial Relations Center, and Audrey Griggs of the Chicago Police Department. A special note of thanks to Robert Emrich, Ph.D., and Louis A. Mayo, both of the Office of Law Enforcement Assistance, U. S. Department of Justice, for their suggestions which added greatly to the value of the project.

TABLE OF CONTENTS

Page

I. INTRODUCTION 1

Need for This Study
Goals of the Present Research
Methods and Procedures
References

II. ANALYSIS OF PATROLMAN WORKING ENVIRONMENT 7

Occupational Analysis
Selection of Districts
Appendix

III. APPRAISAL OF PATROLMAN PERFORMANCE 21

The Paired-Comparison Technique
Background of the Paired-Comparison Rating Technique
Measures Derived from the Paired-Comparison Method
Implementation of the Paired-Comparison Rating in the Districts
Specific Rating Procedures
Call for Volunteers and Selection of Participants
Chicago-Police-Department Performance Measures
Discussion of Eight Criterion Variables
References
Appendix

IV. SELECTION AND DESCRIPTION OF THE PSYCHOLOGICAL TEST BATTERIES 55

Selection of the Tests
Description of the Tests
Appendix

V. RESULTS: PERFORMANCE APPRAISAL 87

Results of the Paired-Comparison Technique
Results of the Call for Volunteers
Results of Chicago-Police-Department Appraisal Measures

Page

VI. VALIDATION OF THE OCCUPATIONAL TEST BATTERIES 97

The Preliminary Multiple-Regression Analysis
The Principal Multiple-Regression Analysis
Identification of Test Variables Most Likely to Predict Patrolman Performance
Prediction of the Paired-Comparison Rating
Prediction of the Chicago-Police-Department Rating
Prediction of Tenure
Identification of the Important Predictor Variables in the Study
Patterns of Test-Variable Scores Likely to Be Most Predictive of Performance
Summary
References

VII. VALIDATION OF AN OCCUPATIONAL TEST BATTERY FOR A RACIALLY-MIXED GROUP 131

Historical Review
Blanket Discrimination
Discrimination against Racial Subgroups and Techniques to Prevent It
Results for Chicago Police Patrolmen: Multiple-Regression Validations
Response of White and Negro Patrolmen to the Test (Predictor) Variables
Response of White and Negro Patrolmen to the Performance (Criterion) Variables
Differential Test Validities for White and Negro Subgroups
Summary and Conclusion
References

VIII. SPECIFIC PATTERNS OF PATROLMAN FIELD PERFORMANCE 187

Method
Discussion
Subgroup 1
Subgroup 2
Subgroup 3
Subgroup 4
Subgroup 5
Subgroup 6
Subgroup 7
Subgroup 8
Conclusion
References

Page

IX. SUMMARY OF RESULTS 223

Goals of the Study
Implementation
Selection of Performance Measures
Selection of Patrolmen Samples
Validation Results
Identification of Patterns of Good and Poor Performance
Implications for Future Research and Application

X. CHICAGO POLICE DEPARTMENT REVIEW OF THE PATROLMAN STUDY 241

History of the Project
Project Accomplishment
Significant Benefits to the Chicago Police Department
Implementation of the Test Battery
Recommendation for Future Project

CHAPTER I

Introduction

This is an age of unprecedented technological advances, in which machines and computers are assuming an increasing share in the work of society. Nevertheless, it is the human resources of an organization that remain the most crucial input in the attainment of that organization's goals. This is particularly true of law-enforcement agencies, given their goal of maintaining justice and order among persons in society. Every man who is part of the "thin blue line" of police is charged with a responsibility so complex and so overwhelming that it must always be entrusted to men and not to their machines.

NEED FOR THIS STUDY

In view of the importance of the law-enforcement function, and the great authority and responsibility vested in the individual police officer, it is strange that behavioral scientists have paid relatively little attention to the manner in which police candidates are selected. Our surprise at the paucity of meaningful research in this area is shared by Blum (1964, pp. 101-102), who makes the following comment about police selection:

> If we restrict the scope of our survey of research to studies which have been made by trained social scientists using acceptable scientific methods for observation and treatment of data, we find that there has been very little such work directed to the evaluation of the usefulness of psychological tests for the selection of policemen.

The apparent lack of scientifically acceptable studies, i.e., studies in which psychological tests or test batteries have been validated for the population on which they will be used, has not deterred police organizations from using tests. Narrol and Levitt (1963) conducted a mail survey which focused on patrolman selection methods of 61 U.S. cities with populations greater than 150,000. Police departments in all these cities were found to use some instrument which could be called a psychological test. In spite of this extensive utilization, the authors concluded that "... there is little or no research being done upon police selection techniques..." and that "not many psychologists are involved either in research upon, or in the use of, the selection procedures now employed."

The use of psychological tests without appropriate validation research is not only totally unwarranted but may be dangerously misleading both to the test user and to the test respondent. Psychologists have long maintained that validation research is a necessity for insuring the proper application of tests. Recently, the federal government moved to support this position through a pamphlet entitled Guidelines on Employment Testing Procedures, issued by the Equal Employment Opportunity Commission (1966). In the document, the Commission states that it adopts the Standards for Educational and Psychological Tests and Manuals, prepared by a joint committee of the American Psychological Association, American Educational Research Association, and National Council on Measurement in Education, and, in essence, makes the validation of tests a prerequisite for their use in selection decisions. The need for selection test validation can no longer be considered a research "luxury"; it is a legal necessity if tests are to be used at all.

GOALS OF THE PRESENT RESEARCH

This brief commentary on the current status of research on police selection clearly demonstrates the need for additional work. In light of this need and in cooperation with the Office of Law Enforcement Assistance and the Chicago Police Department, we planned a comprehensive research study which would deal with two closely related aspects of staffing and strengthening a police organization:

1. The development of effective procedures and the establishment of general standards for the selection of patrolmen.
2. The identification of distinctive "patrolman types," which would exhibit distinctive performance styles in the field and would not be adequately described by the concept of the "average" patrolman.

The first aspect of the study involved a comprehensive validation process to ensure the appropriateness of the chosen tests and their acceptability as tools in a systematic procedure for selecting patrolmen. The second aspect was even more challenging than the first, since it is clear that the concepts of a "typical" or "average" patrolman, or of an "ideal" patrolman, are oversimplifications. A variety of psychological styles may be associated with success or failure, and no single stereotype will sufficiently account for these various styles. We anticipated that the identification of distinctive patrolman types and analysis of their psychological composition and pattern of field performance would provide a rudimentary framework for placement of patrolmen in assignments that would best utilize their skills and behavior patterns.

The study was focused upon a specific population: Chicago Police Department patrolmen, in the Patrol Division, who had at least one year of service and who were currently assigned to uniformed street patrol. This group was thought to be typical of officers employed in large cities throughout the country.

METHODS AND PROCEDURES

The methods and procedures employed in the study, and the specific areas of results, are briefly outlined below. Each heading refers to a subsequent chapter.

Chapter II	Analysis of Patrolman Working Environment--a description of the demands of the field patrolman's job, based on field observation, and a description of the police districts included in the study.

Chapter III	Appraisal of Patrolman Performance--a review of the subjective and objective performance indicators used to characterize individual on-the-job performance, including a special performance appraisal technique, the paired-comparison method.
Chapter IV	Selection and Description of the Psychological Test Batteries--a description of the tests included in the study and of the actual test administration procedures.
Chapter V	Results: Performance Appraisal--a discussion of the results of the implementation of the paired-comparison performance appraisal technique, and the interrelationships of the major and minor performance data obtained from departmental files.
Chapter VI	Validation of the Occupational Test Batteries--a presentation of the results of the multiple-regression analyses for both the primary and the cross-validations, a selection of the areas and tests most predictive of performance, and a description of currently employed patrolmen with respect to the predictive tests.
Chapter VII	Validation of an Occupational Test Battery for a Racially-Mixed Group--a review of the relevant literature, a discussion of strategies to preclude inadvertant discrimination, and a presentation of the test results and validation findings for white and Negro officers.
Chapter VIII	Specific Patterns of Patrolman Field Performance--a presentation of the results of the analysis which identified meaningful performance subgroups within the patrolman population which had been assessed.

Chapter IX	Summary of Results--a complete review of the goals, methods, and objectives achieved, with implications for future research.
Chapter X	Chicago Police Department Review--evaluative comments by Chicago Police Department personnel.

Anyone wishing to read a brief and fairly non-technical review of the highlights of this study should see Chapter IX, the summary.

The Industrial Relations Center of The University of Chicago has been active for almost two decades in manpower-development research and training projects. Although the bulk of the Center's applied research and training has been in industrial organizations and, to a lesser extent, in hospitals and scientific establishments, the past five years have brought an increasing demand for consultation and cooperation in projects in government institutions. The present cooperative research project between the Chicago Police Department and the Industrial Relations Center is a case in point. It has demonstrated, once again, the large areas of common organizational concern shared by such diverse institutional groups as those cited above. More specifically, it illustrates the frequent interchangeability among these organizations of research techniques and training procedures of use in the selection, placement, and utilization of their manpower resources.

This broadly based and far-ranging research project has produced definitive results in many areas, such as the evaluation of departmental procedures for rating patrolman performance, the validation of techniques for predicting patrolman performance and tenure, the identification of specific patterns of performance, and the techniques which should be applied to make fair and equitable selections from among different racial groups. However, caution must be observed in the application of these results. They are specific to the Chicago Police Department. Although they can probably be generalized to other large, urban police departments, their application in any other police department should be preceded by local validations.

This research project was largely exploratory. Its greatest contributions are seen, first, as opening new areas of research and, second, as delineating some of the procedures and identifying some of the techniques and measures most likely to be fruitful in future selection-and-placement validation studies. With regard to the latter, much of the initial exploration, analysis, and fact-finding, which was necessarily a part of the current project, can probably be greatly reduced in local validation studies.

As is often the case with exploratory research in a relatively new domain, more research objectives, both practical and theoretical, suggested themselves than could feasibly be included within the scope of this report. One of these--the selection of candidates from different racial groups--was not a part of the original research proposal but is included here because of its obvious bearing on the results. Others, including refinements in the delineation of the patterns of patrolman performance and an analysis and consolidation of the many predictor variables by factorial methods, are under way at present and will be described in subsidiary reports as time and opportunity permit.

We wish to extend our thanks to the Office of Law Enforcement Assistance for its support, to the Chicago Police Department for its unstinting cooperation, and to the faculty and staff of the Industrial Relations Center, who have directed both interest and enthusiasm to the conduct of this project.

REFERENCES

Blum, Richard H. (Ed.) Police selection. Springfield, Illinois: Charles C. Thomas, Publisher, 1964.

Equal Employment Opportunity Commission. Guidelines on employment testing procedures. Washington, D.C.: EEOC, August 24, 1966.

Narrol, Harvey G. and Levitt, Eugene E. Formal assessment procedures in police selection. Psychological Reports, 1963, 12, 691-694.

CHAPTER II

Analysis of Patrolman Working Environment

OCCUPATIONAL ANALYSIS

The aim of the job-analysis phase of the study was to identify the specific requirements, traits, and personal qualifications necessary for successful performance of the patrolman's job. Rather than attempting to provide an exhaustive list of the many duties performed by patrolmen in the city of Chicago, our scope of inquiry was narrowed to two specific emphases.

Our first real exposure to the organization showed us that the 7,000 patrolmen were by no means a homogeneous group. Rather the great heterogeneity of assignment, ability, and skills which was evident made it mandatory that we limit our focus and consider a group within this total patrolman population. It was decided that the police officers of the Patrol Division who were assigned to beat patrol would exhibit the needed levels of ability and homogeneity of assignment. More importantly, however, the patrol function is an important common denominator across all police organizations, and concentrating our efforts in this common area would provide a greater yield of information to other departments. Thus, our first emphasis was placed upon the activities of the patrolman on the street, rather than on the host of technical, administrative, and specialist functions performed by other uniformed officers in the department.

Our second emphasis in this phase of the study was on the behavioral rather than the "formal" or "official" requirements of the job. A patrolman "enforces law and order," but this phrase lacks real behavioral content. In contrast, stating that a patrolman must "take charge of a situation, e.g., a crime or accident scene, yet not unduly alienate participants or bystanders" describes an element of the job in behavioral terms. Often in the day-to-day activities of the job, the organizational job description is discarded in favor of what the job incumbent "must really do." It was our goal to go beyond the formal job description and capture as well as possible in the time allotted the essential behavioral attributes contributing to success as a police officer.

Method

The job analysis combined "armchair" and field observation to gain as broad and complete an exposure to the police organization and the demands of the police officer's job as time permitted.

Review of published materials and documents provided an important base for a thorough understanding of the job. Departmental and civil service job descriptions (see Appendix) were studied, as were a number of training bulletins and professional journals. Case reports and scenarios of cases which resulted in departmental awards or suspensions were also reviewed. These reviews provided an entree into the "culture" of the police organization and also supplied information for the job analysis.

The field observation began with tours of various facilities in the department, and with meetings and conversations with patrolmen and supervisors at all levels of the organization. However, the most important part of the field observation was the time spent out in the districts and on the street in patrol cars. Four IRC staff members put in many hours on the street observing the patrolmen and acquiring an understanding of the demands of the job. These staff members were exposed to high crime rate, average, and low crime rate districts during all three watches at various times of the week. Being at the district stations and directly observing policemen in their day-to-day activities when riding with patrolmen in beat cars and with supervisors in command cars provided real insight into the behavioral requirements of the patrolman's job. This intensive field exposure was obtained in anticipation of the study and during its early phases, and the experience was available to be utilized in the selection of psychological tests for later application to the patrolman samples. Our field exposure continued throughout the performance evaluation stage in the districts, since the IRC staff conducting the patrolman appraisal were able to go out on the street in many of these districts.

Results

The results of the job analysis are presented below in the form of a list of behavioral requirements. The list is not thought to be all inclusive. Rather it contains many of the attributes which are crucial to the successful performance of the patrolman's job.

It is essential that a patrolman:

1. endure long periods of monotony in routine patrol yet react quickly (almost instantaneously) and effectively to problem situations observed on the street or to orders issued by the radio dispatcher (in much the same way that a combat pilot must react to interception or a target opportunity).
2. gain knowledge of his patrol area, not only of its physical characteristics but also of its normal routine of events and the usual behavior patterns of its residents.
3. exhibit initiative, problem-solving capacity, effective judgment, and imagination in coping with the numerous complex situations he is called upon to face, e.g., a family disturbance, a potential suicide, a robbery in progress, an accident, or a disaster. Police officers themselves clearly recognize this requirement and refer to it as "showing street sense."
4. make prompt and effective decisions, sometimes in life and death situations, and be able to size up a situation quickly and take appropriate action.
5. demonstrate mature judgment, as in deciding whether an arrest is warranted by the circumstances or a warning is sufficient, or in facing a situation where the use of force may be needed.
6. demonstrate critical awareness in discerning signs of out-of-the-ordinary conditions or circumstances which indicate trouble or a crime in progress.

7. exhibit a number of complex psychomotor skills, such as driving a vehicle in normal and emergency situations, firing a weapon accurately under extremely varied conditions, maintaining agility, endurance, and strength, and showing facility in self-defense and apprehension, as in taking a person into custody with a minimum of force.

8. adequately perform the communication and record-keeping functions of the job, including oral reports, preparation of formal case reports, and completion of departmental and court forms.

9. have the facility to act effectively in extremely divergent interpersonal situations. A police officer constantly confronts persons who are acting in violation of the law, ranging from curfew violators to felons. He is constantly confronted by people who are in trouble or who are victims of crimes. Besides his dealings with criminals, he has contact with para-criminals, informers, and people on the border of criminal behavior. (He must also be "alley-wise.") At the same time, he must relate to the people on his beat--businessmen, residents, school officials, visitors, etc. His interpersonal relations must range up and down a continuum defined by friendliness and persuasion on one end and by firmness and force at the other.

10. endure verbal and physical abuse from citizens and offenders (as when placing a person under arrest or facing day-in and day-out race prejudice) while using only necessary force in the performance of his function.

11 exhibit a professional, self-assured presence and a self-confident manner in his conduct when dealing with offenders, the public, and the courts.

12. be capable of restoring equilibrium to social groups, e.g., restoring order in a family fight, in a disagreement between neighbors, or in a clash between rival youth groups.

13. be skillful in questioning suspected offenders, victims, and witnesses of crimes.

14. take charge of situations, e.g., a crime or accident scene, yet not unduly alienate participants or bystanders.

15. be flexible enough to work under loose supervision in most of his day-to-day patrol activities (either alone or as part of a two-man team) and also under the direct supervision of superiors in situations where large numbers of officers are required.

16. tolerate stress in a multitude of forms, such as meeting the violent behavior of a mob, arousing people in a burning building, coping with the pressures of a high-speed chase or a weapon being fired at him, or dealing with a woman bearing a child.

17. exhibit personal courage in the face of dangerous situations which may result in serious injury or death.

18. maintain objectivity while dealing with a host of "special interest" groups, ranging from relatives of offenders to members of the press.

19. maintain a balanced perspective in the face of constant exposure to the worst side of human nature.

20. exhibit a high level of personal integrity and ethical conduct, e.g., refrain from accepting bribes or "favors," provide impartial law enforcement, etc.

Conclusions and Relationship of Occupational Analysis to Test Battery

The chief aim of the occupational analysis was to achieve insight into the behavioral requirements of the patrolman's job so that psychological tests assessing attributes related to these requirements could be included in the test battery administered to the volunteer sample of patrolmen. The analysis supplied information which differed from that usually provided by a formal job description (as well as from a listing of desirable traits, such as "judgment," "tact, " "honesty," etc.). None of the attributes listed above will be a surprise to anyone who has served as a law-enforcement officer, but they provided a quite sufficient basis for selecting of tests for inclusion in the study, as outlined in Chapter IV.

The battery of tests utilized in the study can be classified into the following framework:

Motivational Measures
Measures of Intellectual Ability
Behavioral Measures

Motivational measures were thought to be of paramount importance in the study, since these have apparently found little application in police organizations. In this area, on the assumption that a man's past performance is the best predictor of his future performance, an assessment of the individual's background and experience seemed essential to the prediction of success as a patrolman. In addition, such biographical information could provide data in the complex areas of maturity, ambition, responsibility, etc. Occupational interests and preferences were recognized as another component of motivation, and were also slated for assessment.

A number of the aspects of the patrolman's job indicated that certain mental abilities would be important determinants of success. Reasoning and problem-solving ability were obviously relevant, although the Civil-Service-Commission screening procedure was in part based on an IQ minimum. The ability to comprehend verbal communication seemed important for understanding orders and properly completing administrative forms. Perceptual ability also appeared promising as a predictor, particularly the abilities to perceive differences and details, to organize situations, and to maintain independence of perceptual field.

A test in the area of interpersonal skills was thought to be essential, since a patrolman spends a great deal of his time dealing with others in problematical situations. Therefore, a measure of style of dealing with difficult social situations in a calculated fashion was included in the battery. The items in this measure seemed to parallel much of the routine street contact of the patrolman. The only exception would be extreme emergency situations, which are difficult to generate by means of a test alone.

Because of the many references to personality characteristics in the general job requirements, a number of behavioral tests were also included. It seemed important to include descriptive measures of global behavior patterns such as temperament or needs as well as measures of specific behavioral attributes such as reactivity and stress tolerance.

It was not our intention to attempt a perfect match between a test instrument and each behavioral requirement of the job. Even if this were possible, not all requirements would be equally predictive of successful performance. Using our knowledge of the demands of the job, we therefore attempted to assemble a battery of tests which would broadly cover many of the apparent behavioral requirements identified in the occupational analysis. Further discussion of criteria for selection and an actual description of the instruments utilized are reserved for Chapter IV.

SELECTION OF DISTRICTS

A good deal of attention was devoted to the selection of Chicago Police districts for inclusion in the study, since our goal was a full representation of the diverse environmental elements facing the police officer. A metropolis such as Chicago is, in a sense, patrolled by 21 small police forces, each operating within a fixed geographical area with its own distinctive demographic character. From this point of view, segments of a large city police department may be responsible for policing areas which resemble those patrolled by smaller city and suburban departments. On the other hand, densely populated high crime rate areas are one of the undesirable characteristics of the big city. In any event, an attempt was made to select a wide variety of districts for the study so that patrolmen assigned to a cross section of police work would be participating.

The structure and demographic characteristics of all 21 Chicago Police districts were carefully reviewed in our selection. The definition and demarcation of the districts in Chicago is a complex process, based on a number of factors, such as past workload of each block, natural boundaries (rivers, rail lines, expressways, etc.), special crime problems (arterial streets, public facilities, etc.), and the telephone exchange system within the city (so that calls for police service from within a given district are automatically connected to the radio dispatcher in the Communications Center supervising cars in that district). However, for our purposes, we need only describe the selected districts as they existed at the time of the study.

The geographical area, population density, crime rate, configuration of beats, and manpower of the 21 districts served as the principal objective demographic and structural indices. Districts were selected on the basis of these structural characteristics and of the expert opinion of knowledgeable officials in the department. The districts were also selected to provide a proper representation of

Negro officers in the appraisal and testing phases of the study. A description of the 12 districts selected in terms of the objective indices is presented in Table 1. These districts were thought to reflect the wide spectrum of big city police work, i.e., "fast" districts as well as "slow" districts.

Our aim was to appraise and test patrolmen working in a heterogeneous set of districts. The population of these districts varied from 141,000 to 234,000, and the size ranged from 4.31 to 23.58 square miles. The highest crime rate district in the city was included, as well as some of the "slower" districts. The configuration of beats in the districts was thought to be important. The 2nd District (Wabash), for example, is characterized by beats of small area. This suggests a very high work load. The 5th District (Kensington) has both small and large area beats, suggesting a variety of work-load conditions. The 8th District, on the other hand, is characterized by large area beats which suggests a lower crime rate and large geographical area. (See "Patrol Division Manpower Distribution" included in the Appendix.)

Six of the districts were scheduled for inclusion in Wave I (the first sample) of the study, namely, districts 2, 5, 10, 13, 15, and 19, and the remaining six districts (3, 7, 8, 9, 14, and 21) were included in Wave II. It was not possible at the time of the study to directly quantify the work demand of a particular district for inclusion in a performance prediction model. Individual ability, motivation, and skill are only partial determinants of field performance (although important), and additional factors, such as work load, formal leadership structure, and informal group structure, would ideally be entered in an all-inclusive performance prediction equation. Such an analysis was beyond the scope of the present investigation. A district-by-district analysis of appraisal and test results was not undertaken. Such an analysis may produce additional significant findings, and is contemplated as a worthwhile future project.

CHAPTER II--Table 1

Chicago Police Districts Selected for Inclusion in the Study in Terms of Their Principal Demographic and Structural Characteristics

District		Population of District*		Area of District*		Total Index Crimes***				Crime per Capita	No. of Beats#	No. of Sworn Personnel#
		No.	Rank/21	Sq. Mi.	Rank/21	1964	Rank/21	1965	Rank/21	Rank/21		
2	Wabash	154,831	14	4.31	19	9873	1	8438	1	2	30	469
3	Grand Crossing	179,628	7	5.90	14	7070	5	7099	3	9	27	436
5	Kensington	173,388	9	20.04	4	4425	14	4393	12	12	17	231
7	Englewood	154,806	15	6.53	12	6677	7	6276	5	8	27	407
8	Chicago Lawn	234,374	2	23.58	3	3690	17	3750	14	20	15	200
9	Deering	174,758	8	13.16	6	4596	13	3829	13	13	20	250
10	Marquette	169,549	11	7.24	11	6574	9	5753	8	10	24	347
13	Wood	141,077	16	5.15	16	6613	8	5812	7	7	24	307
14	Shakespeare	181,293	6	7.75	10	4238	16	3535	16	16	20	226
15	Austin	196,747	5	11.95	7	4312	15	3703	15	18	18	214
19	Town Hall	202,484	4	5.70	15	5800	11	4895	11	11	24	254
21	Prairie	128,135	18	4.87	17	7000	6	5282	9	6	18	309

*Source: Chicago Police Statistical Report, 1965.

**The index crimes (murder, rape, robbery, aggravated assault, burglary, theft--$50 and over, and auto theft) are those Part I offenses selected by the F.B.I. for use in making comparisons among jurisdictions.

#Source: CPD, 1966.

CITY OF CHICAGO

Civil Service Commission

announces an examination for

PATROLMAN
Grade P1 - Original

Salary Range: $7,128 - $9,000 per year

Date of Examination: An examination will be held semi-monthly until further notice. Applications will be accepted through Friday afternoon for the examinations to be held on the second and fourth Saturday of each month, starting August 10, 1968

Scope of Examination: Written Examination 100%

Fee: $3.00

Duties: A Patrolman enforces law and maintains order. Protects life and property and assists in the development of proper community attitudes and behavior related to crime, delinquency, prejudice, and civic responsibility. Patrols assigned area by foot or vehicle. Investigates accidents, unusual incidents, and law violations which he observes or which are reported to him. Apprehends persons committing crimes or wanted for investigation concerning law violations. Serves warrants, escorts prisoners to court and places of detention, and testifies in court. May direct pedestrian and vehicular traffic. Performs related duties as required.

NOTE: Candidates may take the Patrolman's examination only once within any 3 month period.

#1030 -- 7/24/68 (over)

Minimum Qualifications:

Applicants must possess weight and measurements as prescribed in Rule III, Section 9, Commission Rules and Regulations, reprinted below and must be physically qualified to meet standards as determined by the medical and physical examiners appointed by the Commission, including these standards of vision:
. corrected vision in both eyes combined must be 20/20
. uncorrected vision in each eye must be at least 20/50

	Weight			Weight	
Height	Min.	Max.	Height	Min.	Max.
5'7"	145	190	6'1"	165	215
5'8"	145	190	6'2"	170	225
5'9"	145	190	6'3"	175	230
5'10"	150	195	6'4"	180	235
5'11"	155	205	6'5"	185	235
6'	160	210	6'6"	190	235

Candidates for this examination must be of good moral character and between 20 and 31 years of age. They must have reached their 20th birthday by the date of examination and must not have reached their 31st birthday by date of examination.

Persons entitled to military preference under Section 10-1-16 of the Illinois Municipal Code (those who have served between September 16, 1940 and July 25, 1947, or at any time between June 25, 1950 and January 31, 1955, or since January 1, 1961, and who were honorably discharged) must not have reached their 35th birthday. Veterans 31 years or over must submit copies of military discharge papers showing dates of service at the time of application.

Applicants for the above original entrance examination must, on the date set for such examination, be citizens of the United States. Applicants must possess a valid Illinois Driver's License at the time of certification. Residency requirements for this examination are waived. Successful candidates will be expected to take up residence in the City of Chicago within a reasonable period after appointment as Patrolman pursuant to the provisions of Section 25-30 of the Municipal Code of Chicago.

APPLICATIONS:

In accordance with the provisions of the Civil Service Act, a fee is charged by the Civil Service Commission for all examinations. Applicants must purchase an Examination Fee Stamp in the office of the City Collector, Room 107, City Hall, to be attached to their applications at the time they are filed. No refund of fees will be made. Commission office hours are 8:30-4:30, weekdays; 9:00-12:00 Noon, Saturdays. The Commission reserves the right in any section of the above examination to impose written, oral, and performance tests and tests of physical qualifications and health.

BY ORDER OF THE COMMISSION:

#1030 -- 7/24/68

JAMES S. OSBORNE, SECRETARY

CHAPTER II--APPENDIX

CODE: 9161
Police Service Group
Police General Duty Series

CLASS TITLE: Patrolman

CHARACTERISTICS OF THE CLASS: Under general supervision, enforces state laws and city ordinances, maintains order, prevents crime, makes arrests; and does related work as required.

EXAMPLES OF DUTIES: Patrols a specified beat or district on foot, on a motorcycle, or in a radio cruiser; checks doors and windows of business establishments, and investigates any suspicious conditions, activities, or persons; makes arrests for violations of laws and ordinances, escorts prisoners to police station, has violators booked on charges, escorts prisoners to court.

Watches for and makes investigations of wanted and missing persons and stolen property; enforces traffic laws, directs traffic and gives information concerning the location of streets, routes and buildings; checks cars for overtime parking, checks parking meters and investigates and makes detailed reports of traffic accidents.

Serves warrants, subpoenas, writs and other court orders as directed and escorts bank messengers and payroll clerks; may supervise clerical personnel performing duties in the police department, perform specialized technical clerical duties such as custody of court evidence, preparation of warrants and summons and instruction work on the firing range with recruits; testifies in court against violators.

DESIRABLE QUALIFICATIONS: Training and Experience. Graduation from high school, preferably with special training in police work; or an equivalent combination of training and experience.

Knowledge, Abilities and Skill. Working knowledge of locations of streets and routes in the City of Chicago; working knowledge of the rights of prisoners and regulations and mechanics of making arrests; working knowledge of departmental rules and regulations.

Ability to observe situations analytically and objectively and to record them clearly and completely; ability to react quickly and calmly in emergencies; ability to express oneself clearly, concisely, orally and in writing; ability to handle situations firmly, courteously, tactfully and impartially; ability to understand and carry out oral and written instructions; ability to develop skill in the use of firearms and in the operation of motor vehicles; ability to use good judgment in evaluating situations and making decisions; ability to perform work involving the necessity for good physical condition and strength.

Elementary skill in the application of modern practices and techniques involving the investigation, arrest, detention and treatment of prisoners, and in the care and use of appropriate tools, equipment and facilities.

CHAPTER II--APPENDIX

CHICAGO POLICE DEPARTMENT
PATROL DIVISION MANPOWER DISTRIBUTION

In 1962, a system was developed whereby Department manpower resources were allocated to the reorganized Patrol Division in sufficient number to provide approximately four hours for preventive patrol to each unit per tour of duty. The experience factors that were analyzed indicated that the average call for service required approximately one hour of patrol unit time. Initially, computations were completed for the third watch (1600 – 2400 hours), since it had the heaviest workload of the three watches and required the maximum number of beats. This determined the beat basis for the Communications Center console maps.

Manpower needs are based on the three watch totals. There are generally fewer beats on the first watch (0000 – 0800 hours) and the second watch (0800 – 1600 hours) than there are on the third watch, and some beats on these two watches are formed by combining third watch beats. The 1st (Central) District is an exception because of the 300,000 people who enter the "Loop" each business day. This district, on weekdays, has a larger number of beats on the second watch. The total number of calls for service during last year's winter or summer season are reduced to a daily average. This average is divided by 4 (the number of calls which can be serviced by a unit and allow approximately 4 hours for preventive patrol during a tour of duty), and this indicates the total number of beats required by the Patrol Division during a twenty-four hour period.

Semi-annually, beats are evaluated to make adjustments to assure maximum effectiveness of beat coverage for either the summer or winter seasons. Among the considerations studied are:

Does the trend of total calls for service indicate an increased need for units?

How effectively did the number of units in the previous summer or winter period satisfy the needs?

Does experience indicate that an excessive number of units were in service?

What is the total number of men and vehicles available for patrol? Should these totals be increased or decreased?

As a result of this evaluation, the total number of units authorized for the Patrol Division is determined.

Patrol Division motorized beats are distributed throughout the City of Chicago in proportion to the weighted workload for the same season in the preceding year. The following weights are used in this determination:

Part I Crimes	4
Part II Crimes	3
Other Calls for Service	1

To lay out the basic beat structure for the City, every call for service is coded for IBM 1410 electronic computer input as to classification (Part I, Part II, Other Calls for Service), street location, and time of occurrence. These codings are weighted by the 4 – 3 – 1 system enabling the computer to produce a listing of the weighted workload for each of the 15,000 square blocks in the City on the three basic watches, 0000 – 0800 hours, 0800 – 1600 hours, and 1600 – 2400 hours.

The police districts are apportioned a percentage of the patrol force by relating their percentage of the city-wide weighted workload to the total number of units available for the City. To add to the validity of the allocation of units determined by weighted workload factors, consideration is also given to the number of calls received by each district. The percentage of the total department calls for service received by each district is compared to the percentage of the department weighted workload in terms of units needed. If this comparison is not reasonably similar, computations are reviewed to identify the cause of the variance.

Also considered is the problem of equating beat size and crime conditions. In the peripheral districts, beats are relatively large while in the core area districts where crime conditions are relatively severe, the beats are quite small. Thus, a patrol unit operating in a very large beat, with the same amount of time available for preventive patrol, cannot be as effective as a unit operating in a smaller beat. In an attempt to equalize police service in large beats, empirical decisions are made authorizing additional patrol units for the peripheral districts when such need is clearly indicated. This process determines the number of patrol beats for each of the 21 districts.

The weighted workload for each district is then divided by the number of beats available for that district to determine the average weight of each beat. The total number of beats required for each watch are determined by applying the percentage of the district workload represented by each watch to the total number of beats available for that district. The patrol beats are laid out to equalize the workload and keep the beat as square as possible while avoiding artificial barriers which would interfere with easy movement throughout the beat. Cover beats are established for the first and second watches so that some vehicles will patrol two beats during hours of reduced activity.

As an example: A district is authorized 54 patrol beats. 41% of the district workload occurred on the third watch, which reflects a need for 22 beats on that watch. 33% of the district workload occurred on the second watch, indicating that 18 beats are required on that watch. 26% of the district workload occurred on the first watch, requiring 14 beats. When the total number of beats in service are reduced, each of the eliminated beats are covered by an adjacent beat car.

Since reported crimes do not coincide exactly with our watch changes, consideration of overlapping watches or "power shifts" is dictated. In some districts, the heaviest workload actually extends beyond the end of the third watch leaving a two and one-half hour period without adequate patrol. This is overcome in the nine districts wherein this condition is apparent by instituting two power shifts. In these districts, the power shifts are superimposed upon the three-watch structure and start at 1030 hours and 1830 hours. By starting the power shifts at this later time, the cars do not leave the street until 0230 hours, thus providing heavier beat coverage as needed.

Conferences are held with District Commanders to determine whether any peculiar patrol needs exist in their districts. In certain districts, it was decided that the power shifts would start at an even later hour on weekends (Friday and Saturday nights), so as to provide heavier coverage further into the early morning hours. Some districts may require power shifts on Friday and Saturday nights only, but not during the week.

The total number of beat units operating on the five watches (3 basic watches and 2 power shifts), are developed from within the total number of units allotted to the district by the Patrol Division.

To provide uninterruped police service to the public during watch changes, roll calls are scheduled at staggered hours.

There is need for both two-man and one-man patrol car operations in developing the patrol pattern for the entire City. From the operating standpoint, advantages are gained by placing one man in each of two squad cars rather than having two men in one car; the size of beats is reduced; the frequency of patrol is increased; and responses to calls for police service are accelerated.

The determination as to which beats will have two-man cars is dependent upon an analysis of types of arrests in the specific area. Special attention is given in this analysis to the number of incidents in which resistance is offered the police; the number in which there are multiple arrestees; the number in which a deadly weapon is involved; and the geographical factors affecting ready access by neighboring squad cars to the area.

The computer produces a priority list, considering these factors, showing the two-man beats in descending order of hazard – the most hazardous listed first. These beats are discussed with the District Commanders so that local factors may also be considered before final decisions are made relative to two-man and one-man cars. After priority of need for two-man patrol beats has been identified, two-man cars are allocated to the extent possible consistent with available manpower.

Prepared by the Planning Division, 29 December 1965.

CHAPTER III

Appraisal of Patrolman Performance

An appraisal of job performance is difficult to obtain on any job, and particularly difficult to secure in assessing the field performance of police officers. The patrolman very often works alone, and receives a minimum of supervision from his Sergeant and Field Lieutenant. These supervisors' knowledge of an officer's field performance derives primarily from their contact with him as he responds to calls for service, and from their reviews of the field case reports he prepares for his assignments during the watch. This supervisory situation is almost the direct opposite of that found in the factory or office, where the subordinate is usually in close proximity throughout the day, and where the supervisor has ample opportunity to observe and discuss the conduct of the work on the spot. In addition, the scheduling of work days and assignments for patrolmen and supervisors makes it unlikely that most men in the district will be supervised by the same Sergeant for their five working days of the week. However, in spite of these apparent difficulties, we believed that the supervisory rating would constitute the single most accurate source of information on an individual's performance. A number of "subjective" and "objective" performance indices were utilized in the study, but the assessment of patrolman performance by the immediate supervisors in the district was given greatest credence.

THE PAIRED-COMPARISON TECHNIQUE

It has been the experience of the Industrial Relations Center that very few industrial organizations are completely satisfied with their employee appraisal procedure. Companies generally have welcomed suggestions for improving their performance review systems. Actually, the Chicago Police Department has a quite adequate system of performance evaluation. However, in planning this phase of the study, it was decided to apply an appraisal system originally developed by the IRC for use in industry. Three reasons supported this decision:

1. There was a need to determine whether the paired-comparison rating system, developed for use in non-law-enforcement organizations, was appropriate for the appraisal of patrolman performance.
2. To insure as accurate a performance evaluation as possible, it seemed desirable to use an appraisal system which had no relevance to the administrative functioning of the department. The results would be used only for purposes of the study, would not be reviewed by senior officers in the district, and would have no bearing on promotions or standing on a promotional list.
3. The comparison of the results of the paired-comparison rating with the semi-annual CPD administrative rating would provide a check on the validity of the CPD rating procedure, as well as confirming or rejecting the appropriateness of the paired-comparison appraisal system for police use.

BACKGROUND OF THE PAIRED-COMPARISON RATING TECHNIQUE*

The paired-comparison rating system has been used by the IRC in a number of industrial and governmental organizations. The method has its origins in the classical experiments conducted by Weber and Fechner about a century ago in the then-new science of psychophysics. Since that time, many psychologists have contributed in adapting this method to the measurement of "psychological" or subjective continua, culminating in L. L. Thurstone's formulation of the laws of comparative judgment. For the specific purpose of appraising job performance, this method requires that every individual to be assessed be compared with every other individual, and an explicit judgment made as to who is the better performer on the job. The number of comparative judgments that would be required if six individuals were assessed is given in Figure 1. The rater would choose the individual in each pair who was the better performer. However, the judgments would not be presented to the assessing supervisor serially as indicated in the figure, but rather in a random order.

---

*Much of the material in this and the following sections has been adapted from Baehr (1963, 1968).

1 vs. 2	2 vs. 3	3 vs. 4	4 vs. 5	5 vs. 6
1 vs. 3	2 vs. 4	3 vs. 5	4 vs. 6	
1 vs. 4	2 vs. 5	3 vs. 6		
1 vs. 5	2 vs. 6			
1 vs. 6				

Total = $\frac{6 \times 5}{2}$ judgments

Figure 1. Comparative Judgments Required by the Method of Paired Comparisons in Assessing Performance of Six Employees

While the paired-comparison method has been generally recognized as yielding the most reliable and valid assessments of any of the contemporary appraisal techniques, it has not been widely used because of the time and effort required by the rater in handling large numbers of pairs and because of the computational labor involved in obtaining the scale values. This is especially true when there are a sizable number of individuals to be assessed, since the number of comparisons rises as an exponential function $[\frac{n\,(n-1)}{2}]$ of the number of individuals (n) being assessed.

Because of this problem and the problem of obtaining scale values, the history of the use of the paired-comparison method has consisted largely of successive innovations in the mechanics of administration and scoring. A device called The Comparator (1961, 1967) has been developed for appraising the performance of up to 19 individuals. However, the size of the watches in the Chicago Police districts required the use of more recently designed computer-based techniques.

The Industrial Relations Center has developed two interconnected computer programs to implement the paired-comparison rating procedure in larger groups. The first, called PAIRS, is designed to pair every individual in the group being appraised with every other individual in the group. Each pair of names is printed on a separate IBM card, and the order of the names in the deck as a whole is randomized. The prepared decks of IBM cards are given to the raters with instructions to place a check mark in a box adjacent to the name of the individual in each pair who is doing the better job. The scoring of the completed IBM deck is done by another computer program called COMPARE. Surprisingly little computer time is necessary to produce the scale measures to be discussed shortly, even in large-scale multi-judge appraisals.

MEASURES DERIVED FROM THE PAIRED-COMPARISON METHOD

Although a variety of measures can be derived from the paired-comparison method, three were particularly germane to this study:

1. The per cent of consistent judgments for each rater.
2. The per cent of agreement between or among raters.
3. The performance index for each patrolman.

The per-cent-consistent-judgment score, which ranges from 0 to 100 per cent, represents the consistency of judgments made by each individual rater. The concept of consistency of judgments is illustrated in Figure 2. For example, if a rater selects patrolman A over patrolman B, and patrolman B over patrolman C, then, to be consistent, he should also select A over C. If he selects C over A, this choice is regarded as an inconsistency. For any set of "n" subjects, there is a maximum possible number of inconsistencies. The per cent of inconsistent judgments is the number of inconsistencies actually found expressed as a percentage of the number of possible inconsistencies. This percentage is subtracted from 100 per cent in order to obtain a positively stated "per cent of consistent judgments."

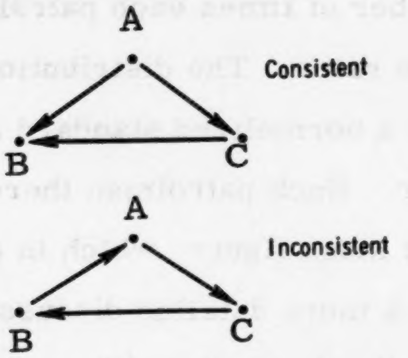


Figure 2. Consistency of Judgments

The per cent of agreement between or among raters applies only when two or more judges have rated the same group of individuals or a common group of individuals within two otherwise dissimilar groups. The per cent of agreement is the ratio (expressed as a percentage) of the number of pairs of similar judgments to the total number of judgments made concerning the individuals being rated in common. With two raters, this score ranges from 0 to 100 per cent, and 100 per cent always indicates perfect agreement no matter how many raters are involved. However, the lower limit of the range varies according to the number of raters contributing to the coefficient.

It should be noted that on the basis of the two indices described above, it was possible to exclude ratings from those raters who were either internally inconsistent or in marked disagreement with other raters in the district. It was also possible to exclude those patrolmen on whom the raters most disagreed--usually subjects who had only recently joined the department or who were relatively little known to some of the raters. Such exclusions can increase the validity of the third measure derived from the paired-comparison method--the performance index derived from the ratings, which is described on the next page.

The basic information that was needed to calculate the performance index was the number of times each patrolman in the group was chosen over all others by the rater. The distribution of each rater's preference was converted to a normalized standard score scale by direct calculation on the computer. Each patrolman therefore received a standard score performance index figure, which in practice varies between a low of 30 and a high of 70. A more detailed discussion of normalized standard scores is presented in the Appendix. Where two or more raters evaluated the performance of a patrolman, their standard score ratings were averaged to provide a single composite performance index.

IMPLEMENTATION OF THE PAIRED-COMPARISON RATING IN THE DISTRICTS

The paired-comparison performance rating was conducted by members of the IRC staff in one district at a time in each of the districts included in Wave I and Wave II. The usual sequence of events began with a meeting in the office of the District Commander, at which time the Director of Personnel and his Assistant provided a general orientation to the study and introduced the IRC staff members who would be conducting the appraisal. The Field Director of the study then provided a more specific outline and discussed in detail the requirements of the performance appraisal phase. The objective of these meetings was threefold--to acquaint the District Commander with the study, elicit his cooperation, and identify qualified raters within the district.

It was decided to limit the choice of raters to Supervisory Sergeants and Field Lieutenants who had at least three months tenure in the district and who were well acquainted with the performance of at least ten patrolmen. Limiting the raters to the two above mentioned ranks excluded a number of Sergeants who worked exclusively in administrative capacities, the Captains who generally functioned as Watch Commanders, and the District Commander himself. The bulk of the rating responsibility was therefore placed on supervisors most likely to have had ample opportunity to directly observe a man's field performance. The time limit and the minimum number of patrolmen the rater had to be acquainted with excluded supervisors who had not had sufficient opportunity to gain knowledge of the performance of patrolmen in the district.

During the performance appraisal in the Wave I districts, supervisors were individually contacted by IRC staff and given a short briefing on the study. (Private office space in the district station was provided by the District Commander.) This briefing concentrated upon five points:

1. The study was sponsored by the Office of Law Enforcement Assistance, U. S. Department of Justice.
2. The objective was to identify tests which might be useful in selecting patrolmen.
3. The supervisory rating was to be the foundation of the study and was to provide a meaningful standard of patrolman performance to use in deciding which tests had value and which did not.
4. The ratings were for research purposes only, were strictly confidential, and were the property of the IRC. The results of the rating would not be put into any man's personnel folder and would in no way influence a patrolman's career in the Department. The supervisory ratings would not be seen by anyone in the Department, and the supervisor's own identity as a rater would remain confidential.
5. The actual rating procedure to be followed at the second meeting--when the IBM rating cards based on the supervisor's rating list had been prepared--was described.

After answering any questions the supervisor had, the IRC staff member presented him with a roster of the watch, and asked him to indicate which men he felt he know well enough to rate, with the following provisos:

1. Select only patrolmen whose current field performance is known to you.

2. Do not select patrolmen who have been in the district or on the watch too short a time for proper evaluation.
3. Select only "outside" men, i.e., patrolmen assigned to street duty. (This requirement excluded patrolmen assigned to desk, lock-up, review office, or other administration duties.)
4. Do not, however, include "outside" men who are not, or have not recently been, assigned to uniformed street patrol. (This requirement excluded patrolmen who served only on intersection control or public building assignments, and patrolmen assigned plainclothes duty as Vice-Dectectives in the district. Members of the District Tactical Unit, a special unit organized in the districts during the course of the study, were ordinarily included, since most of these men entered this assignment from uniformed patrol duty.)

Specifically the patrolmen included were officers assigned to uniformed patrol in the vehicles of the district with sufficient tenure to allow for an informed supervisory evaluation.

The supervisors were encouraged to include as many patrolmen as they knew well enough to rate, with a desired minimum of 20 and a desired maximum of 40 officers. A point of clarification occasionally provided was that raters should not limit their choices to outstanding patrolmen, but should rather attempt to provide a list which contained a full representation of field performance--superior, average, and below-average patrolmen. Once the lists of names for rating were completed, arrangements were made to meet with the supervisor when the IBM rating deck was prepared.

During Wave II, the introductory meetings with supervisors in the districts were conducted somewhat differently. A great deal of time was spent during Wave I in providing the basic orientation to supervisors. While arrangements were being made for implementing the appraisal in Wave II districts, the Commander of the Third District suggested that we make the orientation and development of rating lists part of his regular monthly staff meeting. This procedure enabled us to contact the majority of the supervisors in the district in a single visit and greatly facilitated communication about the study as well as preparation of the rating materials. The orientation remained the same, except that there were 20 to 25 supervisors present rather than one or two. All districts in Wave II cooperated in making time available during their command staff meetings for presenting the study objectives and obtaining supervisors' rating lists.

SPECIFIC RATING PROCEDURES

A deck of IBM rating cards was prepared for each supervisor in the district, based on the roster of names obtained from him in the first meeting with the IRC field worker. As described above, this deck contained a pair-by-pair comparison of all possible combinations of patrolmen on his list in random order. Each man on the list was therefore paired with every other man in the deck. The supervisor was given the following instructions:

1. This deck of IBM cards is based on the roster obtained in our first meeting. Each card contains the names of two men, and the rating deck consists of all possible pairs of names on your list.
2. Try not to consider what you may have heard others say about these men. Rather, base your judgments upon your own experience with them.
3. Consider the two men on each card, and ask yourself the following question: "Which of these two men is the better performer on the street--which is the better patrolman in terms of performance in the field?"
4. When you have made your choice, place a check mark in the box next to the name of the man that you feel is doing the better job. Make a choice for each card and make only one check mark on each card.

5. This is the only judgment you have to make. You do not have to rate any specific characteristics or traits of these officers, nor do you have to justify the choices that you make. All you do is decide which of the two men on each card is the better patrolman, and then place a check mark next to his name. Your choice should be based on (1) the factors you consider important for effective performance as a patrolman, using standards you have developed as a police officer and as a police supervisor and (2) your knowledge of the job performance of each man.
6. Work rapidly, and do not spend too much time on any one card. The rating of a deck of cards based on a list of 30 names can be completed in approximately 30 minutes.

The above instructions were given to all supervisors participating in the study. The rating was generally conducted in a private office at the district station on an individual basis. It was occasionally possible to have two or more raters work at the same time on their rating decks. In these instances, each rater's judgments were made independently of the others, and no discussion of any particular individual was permitted. The results of the paired-comparison performance appraisal are given in Chapter V.

CALL FOR VOLUNTEERS AND SELECTION OF PARTICIPANTS

Several weeks before the scheduled group testing sessions, IRC and Personnel Division staff met with each of the District Commanders. The Field Director reviewed the experiences and progress of the IRC staff in conducting the performance appraisals in the districts, and provided a detailed description of the testing phase of the study and the proposed "ground rules" for patrolman participation. Early in the study it had been decided that calling for paid volunteers to participate on their own time would provide a more meaningful estimate of patrolman ability than ordering a number of men to take part. It would, of course, have been desirable to have all the patrolmen in a district volunteer, but this was judged unlikely because of secondary employment, vacations, and mere disinclination to participate.

District Commanders were asked to announce this phase of the study at roll-call meetings (held at the district before the watch went out on patrol), distribute participation forms (Appendix), and request that these forms be completed and turned in. The IRC provided each District Commander with an outline for the announcement containing a description of the study in sufficient detail to answer questions which might be raised by the patrolmen. A copy of this outline is presented in the Appendix.

It should be pointed out that, in addition to the press coverage of the study at the time of the grant award, patrolmen in the department were exposed to news of the study as early as October of 1966, when an article describing the project appeared in the Chicago Police Star magazine (Appendix), well in advance of the Wave I (February, 1967) and the Wave II (July, 1967) testing. In addition, a notice of the study appeared in the Daily Bulletin at the same time the roll-call announcements were made at the districts (Appendix).

A deadline was set for the return of the Patrolman Participation Forms, and in most cases the District Commander's secretary was responsible for collecting them. It is important to note that we did not pressure the District Commander to obtain a 100 per cent volunteer response. We felt that undue pressure to participate would be just as deleterious to meaningful test performance as ordering men to participate.

Forms were picked up at the district stations and processed at the IRC. Volunteers were separated from non-volunteers, and were considered for participation in the light of three elements of information derived from the performance appraisal results:

1. Evaluation by at least two supervisors in the district.
2. Acceptable levels of individual consistency and agreement in the

pooled supervisory ratings. In practice, this meant that raters were at least 90 per cent consistent, and at least in 60 per cent agreement in rating the performance of the individual patrolman being considered.

3. A performance index less than or equal to 46 (the bottom third of performance) or greater than or equal to 54 (the top third of performance).

Individuals meeting these criteria were selected for inclusion in the study, scheduled for their preferred location and test session, and notified of their selection by means of the Patrolman Confirmation Card (Appendix).

CHICAGO POLICE DEPARTMENT PERFORMANCE MEASURES

An extensive review of sources of possible information on patrolman performance available within the Chicago Police Department was made by a joint CPD-IRC committee. Approximately 160 elements of information were selected for consideration from three principal categories:

1. Civil Service Commission and Background Data
2. Police Training Academy Record
3. Patrolman Performance Data (past and current)

In addition, a number of elements of descriptive data (e.g., star number, race, etc.) were available.

All selected elements were thought to have some relevance for assessing the current performance of Chicago patrolmen, but the list was cut in half by applying three criteria:

1. Accuracy: Some question existed as to the accuracy of some of the information on file. Marital status best illustrates this problem, since this information was obtained at the time of recruitment, and changes via marriage or divorce were not necessarily reported by the individual officer.
2. Meaningfulness: In the case of civil service scores, for example,

data on tests used were sometimes incomplete and thus raw scores were uninformative since it was not known how many items they were based on.

3. Availability. This was easily the most crucial criterion. Information did not exist for significant portions of the sample, since record-keeping policies and procedures were simply altered over the years. For example, the Watson-Glaser Test of Critical Thinking is now routinely administered in the Training Academy at the present time. However, it is not available for a number of patrolmen who were tested merely because it was not administered during their Academy training. The retrievability of information was also considered, i.e., the cost of obtaining a particular element of information in relation to its potential value in the study. Lastly, a certain amount of data was found to be missing, through filing errors, omissions, or the like.

Approximately 80 elements of detailed performance information on each patrolman tested were entered into his IRC data file. This information was provided by the Personnel Division upon receipt of the roster of patrolmen tested, and in no way influenced the selection of participants. The data fell into the following areas:

1. Educational preparation
2. Prior employment (civilian and military)
3. Training Academy ratings and test scores
4. Absence and injury record
5. Internal Investigation Division and credit complaints
6. Performance rating record (1963 through 1966)
7. Departmental awards
8. Disciplinary actions
9. Arrest record (number and type of arrests made in 1966)

Not all of these performance measures seemed equally important. Therefore, the IRC staff attempted to further refine this list to arrive at the measures which would be most suitable for use as criteria in the multiple-regression analysis. Our refinement was based upon substantive as well as technical grounds. Items in the final listing, presented in Figure 3, were thought to provide a comprehensive picture of patrolman performance as assessed in a number of distinct ways. Technical considerations were also important, for example, the nature of the distributions of each variable. Missing data played an important role, since a case would be deleted from the study if any of the test or performance information was lacking.

DISCUSSION OF EIGHT CRITERION VARIABLES

The paired-comparison rating, which yielded a composite performance index for each patrolman included in the study, was one of the principal criterion variables. The derivation of this measure as well as its characteristics have been thoroughly discussed in this chapter, and will not be commented upon further.

The Chicago-Police-Department performance rating was the second principal criterion variable. This measure was based on the average of the two total scores provided by district supervisors in the two semi-annual ratings (January-June and July-December) conducted during 1966. The Appendix contains a Department Special Order which describes the rating procedure, a performance rating checklist developed to assist the supervisor in preparing his ratings, and the actual performance rating card. This rating was available for all patrolmen tested, and was thought to provide a somewhat different type of information than that obtained in the paired-comparison rating. Both rating procedures were aimed at obtaining a global index of patrolman performance, but while the paired-comparison rating emphasized the field performance of the officer, the CPD rating emphasized a more conventional delineation of job performance in terms of quality and quantity of work, dependability, personal relationships, and attendance and promptness. All these elements were given equal weight in determining the CPD rating, while no specific categorization or weighting of traits was

CHAPTER III -- Figure 3

CRITERION VARIABLES USED FOR MULTIPLE REGRESSION ANALYSES

	MEASURE	DESCRIPTION
1.	Paired-Comparison Rating	From the performance appraisal conducted by the IRC staff
2.	1966 CPD Performance Rating	The average of the two half year totals from the CPD
3.	CPD Tenure	Years of service on the CPD
4.	CPD Total Awards	Honorable mentions + commendations + higher awards averaged for 1961 through 1966
5.	CPD IID Complaints	Ratio of IID complaints over IID sustained complaints
6.	Disciplinary Actions	Disciplinary actions averaged for 1961 through 1966 + IID sustained complaints
7.	Number of Arrests, 1966	Total arrest record for 1966
8.	Times Absent	Total incidence of absence for 1966

performed in the paired-comparison rating. Essentially, the paired-comparison rating focused on field performance, while the CPD rating emphasized the administrative as well as field performance aspects of the job.

The third criterion variable was tenure in the Chicago Police Department. The use of years of service as a standard of performance was important from the point of view of understanding patrolman turnover. It was also promising for understanding the stability of employment and the likelihood of an individual remaining a police officer and serving until retirement.

The awards received by a patrolman during his tenure were thought to be another important source of information on his performance. The CPD total-awards index was based upon the average number of complimentary letters, honorable mentions, commendations, and higher awards received from the department in the period 1961 through 1966. The index for patrolmen with less than six years of tenure was based upon their own length of service.

A review of complaints registered with the Internal Investigation Division (IID) of the department led to the establishment of the fifth criterion measure. This index was based upon the ratio of complaints over sustained complaints registered in the IID during the patrolman's tenure with the department.

The number of oral and written disciplinary actions received by a patrolman were averaged over his years of service (up to and including the last six years) and added to the number of sustained IID complaints incurred to consitute the disciplinary-actions index.

The Data Systems Division of the department cooperated in making available the arrest performance of each officer included in the study. The figure used for the number-of-arrests criterion variable was the total of the number of arrests in all categories made during 1966.

The eighth criterion variable was the attendance record of the officer during 1966. The total incidence of absence in this year constituted the times-absent index.

The eight major criterion variables were thought to provide multi-faceted information on patrolman performance. The indices were both meaningful and useful when used as standards of performance, and were considered appropriate for use as criteria in the multiple-regression analyses (Chapters VI and VII). These performance measures were also used as the basis for defining performance subgroups within the general patrolman sample (Chapter VIII).

REFERENCES

Baehr, Melany E. The appraisal of job performance (with special reference to the use of the Comparator). (Occasional Paper Number 27) Chicago, Ill.: Industrial Relations Center, The University of Chicago, 1963.

Baehr, Melany E. The appraisal of job performance (results of five years' use of the paired-comparison technique for the validation of selection test batteries). (Occasional Paper Number 27-R1) Chicago, Ill.: Industrial Relations Center, The University of Chicago, 1968. (In press)

Tierney, John B. and Biggins, Peter A. The Comparator (instrument, 1961, and Test Administration Manual--Revised 1967). Chicago, Ill.: Industrial Relations Center, The University of Chicago.

CHAPTER III--APPENDIX

Discussion of Normalized Standard Scores

A commonly-used way of expressing relative standing in a group is the percentile. Job performance and test scores are often reported in percentiles to indicate the relative level at which an individual scores as compared to a larger normative group scoring below a certain level. For example, a percentile score of 68 indicates that 68 per cent of the normative group scored below this level on the measure.

A number of the results in this report, however, are presented in a somewhat different manner--in the form of normalized standard scores. Standard scores are used because of certain statistical advantages they provide in the analysis of the results. For example, they permit the averaging of performance indices based on two or more raters, an operation which would be technically incorrect with percentage figures. Standard scores are easily converted to the more common percentile, for those more familiar with this reporting procedure. The standard score ranges may be described by their equivalent percentile ranges in the following manner:

Standard Score Range	Percentile Range
20-30 (Very Low)	1st to 2nd Percentile
30-40 (Low)	2nd to 16th Percentile
40-45 (Low Average)	16th to 31st Percentile
45-55 (Average)	31st to 69th Percentile
55-60 (High Average)	69th to 84th Percentile
60-70	84th to 98th Percentile
70-80	98th to 99th Percentile

A more detailed conversion table is presented on the following page.

CHAPTER III--APPENDIX

CONVERSION TABLE

FOR CONVERTING NORMALIZED STANDARD SCORES INTO PERCENTILES

To convert normalized standard scores into percentiles, locate the standard score in the column headed Standard Score and read across under the Percentile column to get the equivalent percentile. This table is included for the convenience of those test users accustomed to percentiles.

Standard Score	Percentile	Standard Score	Percentile
72 & higher	99	50	50
71	98	49	46
70	98	48	42
69	97	47	38
68	96	46	34
67	96	45	31
66	95	44	27
65	93	43	24
64	92	42	21
63	90	41	18
62	88	40	16
61	86	39	14
60	84	38	12
59	82	37	10
58	79	36	8
57	76	35	7
56	73	34	5
55	69	33	4
54	66	32	4
53	62	31	3
52	58	30	2
51	54	29	2
50	50	28 & below	1

CHAPTER III--APPENDIX

PATROLMAN PARTICIPATION FORM

(Please Print)

Name ______________________________ Star No. __________

District __________ 3rd Police Period: (2 — 29 March) Watch ________________

Day-Off Key ____________

________ I will participate in the U.S. Department of Justice Study, if selected. I understand that payment of fifteen dollars will be made to selected participants. Participants will be notified during the week of 13 March 1967.

Indicate preferred test session below, by number:
(1 = first preference, 2 = second preference)

________ Session 1, Mon, 20 March 1967 - 0900 hrs Law School, Univ. of Chicago*
________ " 2, Mon, 20 March 1967 - 1830 hrs " "
________ " 3, Wed, 22 March 1967 - 0900 hrs " "
________ " 4, Wed, 22 March 1967 - 1830 hrs " "
________ " 5, Thur, 23 March 1967 - 0900 hrs Hall of Science, DePaul Univ.**
________ " 6, Thur, 23 March 1967 - 1830 hrs " "
________ " 7, Mon, 27 March 1967 - 0900 hrs " "
________ " 8, Mon, 27 March 1967 - 1830 hrs " "

(Selected Participants Will Be Notified the Week of 13 March 1967)

________ I will not participate in the U.S. Department of Justice Study.

Date of Return ________ Signature ______________________________

(To be returned to District Commander not later than 7 March 1967)

*1111 East 60th Street

**1036 West Belden Avenue (Belden - 2300 North)

LEAA Patrolman Study - 2/67.

CHAPTER III--APPENDIX

REVISED OUTLINE

June 19, 1967

To:	District Commanders
From:	John Furcon Field Director, LEAA Patrolman Study
In Re:	Suggested Outline for Roll Call Announcement of LEAA Study of Patrolmen and Call for Volunteers

1. Background of the Project--This study was proposed by the Industrial Relations Center of The University of Chicago, and, with the sponsorship of the Chicago Police Department, was submitted to the U.S. Department of Justice review board. The study was made possible by a $40,000 grant through the Law Enforcement Assistance Act of 1965, which provides for research projects which would improve law enforcement.

2. Purpose of the Project--The purpose of the study is to identify characteristics or attributes which are related to success as a patrolman. The project essentially explores the following questions: What are the qualifications necessary to meet the requirements of the patrolman's job? What are the different "types" of patrolmen--in what ways are patrolmen alike, and in what ways are they different? How are the characteristics of these similar groups related to their assignments and performance?

3. Progress to Date--The project began in August, 1966, with a field study of the patrolman's job. Observations were made by "riding along" with patrolmen and sergeants, and on the basis of this experience, a number of tests were selected as being appropriate for the police situation. During November and December, a performance appraisal of the patrolmen was made by the command personnel of the Districts involved (2, 5, 10, 13, 15 and 19) for the later selection of volunteers. A call for volunteers was made in these Districts during January, 1967, but due to the blizzard, the testing which was scheduled for February was postponed. A second call for volunteers was made in these Districts during February, and approximately 70% of the patrolmen volunteered to participate. A total of 265 volunteers were selected, and approximately 250 patrolmen completed the series of tests during March, 1967. The analysis of their scores is still in progress.

During April and May, 1967, performance appraisals of patrolmen were made by supervisors in a second group of Districts. A call for volunteers in Districts 3, 7, 8, 9, 14 and 21 is planned for June 22-28, 1967. In addition, a call for volunteers will be made to Task Force Tactical patrolmen.

4. Requirements for Participants--The assistance of patrolmen in the District is crucial to the success of the study. The call for volunteers is being made in this District and in the others listed above, with the hope that an even larger percentage of the patrolmen will volunteer. The number of participants actually selected will be much smaller than the anticipated number of volunteers, since a representative sample of patrolmen in the Department will be selected from among the volunteers on the basis of the supervisory performance evaluation. Approximately 50 men, representing a cross section of patrolmen, will be selected from each district.

   Participants who are selected by the Industrial Relations Center will be requested to complete a series of personnel tests, requiring approximately four hours to complete. These tests will, in general, be different from the usual "I-Q" tests, and will instead focus on areas such as perceptual ability, reaction time, and behavioral characteristics thought to be particularly important for patrolmen. The tests will be administered in group sessions. Some of the tests will have strict time limits; others will allow as much time as needed for completion.

5. Conditions of Participation--Patrolman participation will take place under the following conditions:
   a. All test results are confidential. Individual test booklets and scores will be retained by the Industrial Relations Center.
   b. Individual test results will not be disclosed to the Chicago Police Department. No record of scores will be placed in a man's personnel file.
   c. No administrative action will be based on these results. Since individual scores will not be disclosed, no promotion decisions will be based on them. Test results in this study will in no way help or hinder a man's career.
   d. Individual identity will not be important. Although the testing will not be anonymous, individual identity will be used only to relate test scores to performance evaluation. Findings will be presented in terms of group rather than individual results.

6. Response to Call for Volunteers--Each patrolman in the district will be required to complete a patrolman participation form. If a patrolman decides to volunteer, he will also indicate his preferences among the scheduled test sessions--either morning or evening, North side or South side. Patrolmen who decide not to volunteer will also indicate their response. The participation forms will be collected and returned to the Industrial Relations Center, and the selected volunteers will be promptly notified.

7. Benefits to Be Derived--A number of individual and departmental benefits will be derived from the study:
   a. Participants will receive a stipend of $15 for their cooperation.
   b. The successful completion of the study will have a positive effect on the trend toward professionalization of police work.
   c. The study is very likely to have national significance in that findings established in Chicago may result in changes and improvements in the method of selecting and assigning new patrolmen in all large police departments throughout the country. This national application of results will reflect to the credit of the districts involved, as well as to that of the department as a whole.
   d. The successful completion of the study will provide an opportunity for the Chicago Police Department to take the lead in police personnel research--an area which has been relatively barren until the present time.

TIME SCHEDULE

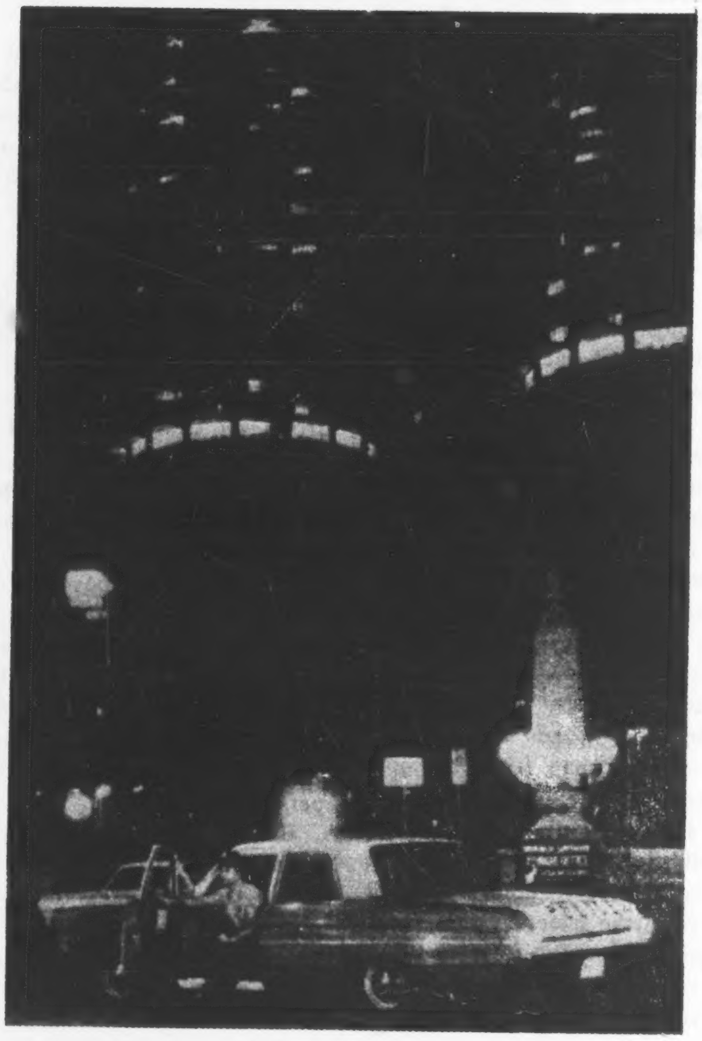
21 June 1967 to 22 June 1967	Receipt of outline and patrolman participation forms at each participating district.
22 June 1967 to 28 June 1967	Announcement and discussions of project at roll calls by District Commanders.
28 June 1967	Patrolman participation forms picked up at district stations by Industrial Relations Center staff.
3 July 1967	Notification of selected participants.
10 July 1967	0900 and 1830 test sessions at University of Chicago, Law School, 1111 E. 60th Street
12 July 1967	0900 and 1830 test sessions at University of Chicago, Law School, 1111 E. 60th Street
14 July 1967	0900 and 1830 test sessions at University of Chicago, Law School, 1111 E. 60th Street
18 July 1967	0900 and 1830 test sessions at De Paul University, Liberal Arts Building, 2322 N. Kenmore

CHAPTER III--APPENDIX

CHICAGO POLICE STAR

OCTOBER, 1966

THE OFFICIAL PUBLICATION OF THE CHICAGO POLICE DEPARTMENT



Pioneer Study of Patrolmen—page 16

Pioneer Study of Patrolmen

THERE IS NO "IDEAL" policeman, no "typical" policeman—and researchers are not looking for him. That statement clarifies one of the misconceptions about a new research study now being made by the University of Chicago's Industrial Relations Center in cooperation with the Police Department. For the first time, researchers are making an intensive study of police patrolmen, using techniques already found to be successful in other government agencies and in industry. It is a pioneer study to determine the requirements of the patrolmen's job and the qualifications needed to meet them.

The study was made possible by a $40,000 grant under the Law Enforcement Assistance Act (LEAA) by the U.S. Department of Justice, Office of Law Enforcement Assistance. The LEAA provides grants of federal funds to law enforcement agencies for projects which would improve law enforcement. This study will eventually make an impact on recruitment and assignment policies in police departments throughout the country.

WHAT IS THE OBJECTIVE OF THE STUDY?

1. The most important objective is to ascertain the characteristics or attributes typical of successful, competent patrolmen. What are the different types of policemen which are equated with success? In what ways are patrolmen like each other, in which ways are they different? And how are these various "sub-groups" within the "police population" related to their present assignments and performances?

2. Once these various groups of successful patrolmen are identified and defined, general standards for officers—far more refined than anything now existing—can be set.

3. The tests that will be used in the project will themselves be "tested". In other words, which tests are most useful and effective in the selection and placement of patrolmen? Which ones don't really tell recruiters and personnel men much at all, which tell them a great deal?

HOW DOES THE STUDY WORK?

A staff made up of members of the Department's Personnel Division and the Industrial Relations Center are directing the study. The University's research team is headed by Dr. Melany E. Baehr, Director of Measurement Research, Industrial Relations Center. Field Director for the project is John E. Furcon, and Ernest C. Froemel is in charge of computer processing. Both men are graduate students now working on their Ph.D.'s at the University, and both also teach at DePaul University.

The testing will be done in two parts. Ten districts in all will be selected. Testing in five of them will start next month; testing in another five will be conducted in the spring. Each district selected will have a good cross-section of communities—from quiet, residential areas to high-crime areas.

Field researchers will go into these districts, interview sergeants and ask them to rate their men. Moreover, researchers will ask two or three other sergeants in the district to rate each man. And in addition, the field lieutenants will rate the men.

After the men are rated, volunteers will be asked to go to the University of Chicago for a half-day session to take a battery of tests. Not all volunteers will be accepted — perhaps 45 to 50 from each district, a total of 400 to 450 for the entire project. The testing will be done during the patrolmen's off-duty hours, and each will be paid $15 for his time.

All testing will be done on a confidential basis. The University of Chicago will give the tests and process the results. None of the participating officers will be individually identified to the Department, and none of the information will ever be placed in the officer's personnel jacket.

Because of the complexity of the patrolmen's job, a variety of tests will be given—aptitude tests for specific skills and specific mental abilities (not just IQ), for temperament, emotional adjustment, interests outside the job, etc.

These tests will be scored, punched on IBM cards and fed into a computer and eventually put on magnetic tapes and discs. Without computers, a study of this magnitude could never be made.

The two studies, in the fall and in the spring, will be analyzed. Then a cross-check will be made to validate the findings. The two studies will then be combined and all information obtained will be analyzed.

WHAT WILL BE THE RESULT OF THIS PROJECT?

There will be no immediate impact—the study is expected to take 16 months, perhaps more, before all data are analyzed and a report written.

Again, it will not have any effect on the individuals who take the tests—not now or at any time in the future will it in any way help or hurt his career, because this information will be kept confidential.

But eventually, it is expected to become of nationwide importance to police departments. The study will help clarify the demands made upon and the variety of tasks required of a patrolman, and the knowledge and skills he must have.

Once the various kinds of successful or competent patrolmen can be identified, and the characteristics and attributes which are typical of them determined, these characteristics and attributes would then be pinpointed by means of recruitment tests. The tests that prove to be most effective in defining the various characteristics that are typical of these "sub-groups" of successful patrolmen would be used as selection tools.

With a better understanding of the job requirements and performance abilities needed for that job, men could be assigned to better utilize their specific abilities, aptitudes and skills. Manpower resources of a department could be used more effectively.

The project is designed so that other police departments throughout the country will be able to use the results. It is another major step forward in the professionalization of policemen. ★

Reprinted with permission
of the
Chicago Police Star, 1966

CHICAGO POLICE DEPARTMENT

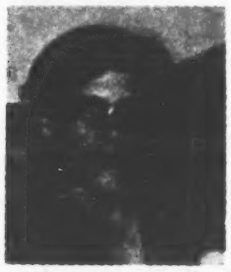
DAILY BULLETIN

VOLUME 8, NUMBER 175
25 June 1967

CONFIDENTIAL – FOR POLICE USE ONLY

O.W. WILSON, SUPERINTENDENT

MISSING PERSON



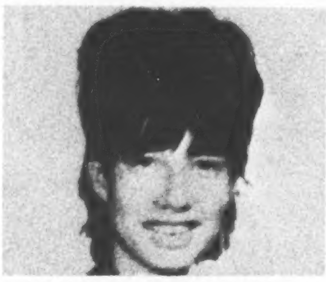
JAMES HEDIN
1711 Berwyn Ave.

M/W, 16, 5-6, 140, med bld, fair comp, hazel eyes, brn hair. Wearing blue sweater with black top, blk trousers, blk shoes, blk jacket, blue shirt. May be in Evanston where father lives or around Amundsen High School, where he formerly attended. RD F071816. Area 6 Youth Sec.

ARE YOU AWARE?

University of Chicago seeking patrolmen for police survey. Desire officers from Districts 3-7-8-9-14-21 and Task Force. Volunteers will be requested at roll call. Volunteers will be paid $15.00 per day. Survey sponsored by U. S. Dept. of Justice and Industrial Relations Cnr., University of Chicago.

MISSING PERSON



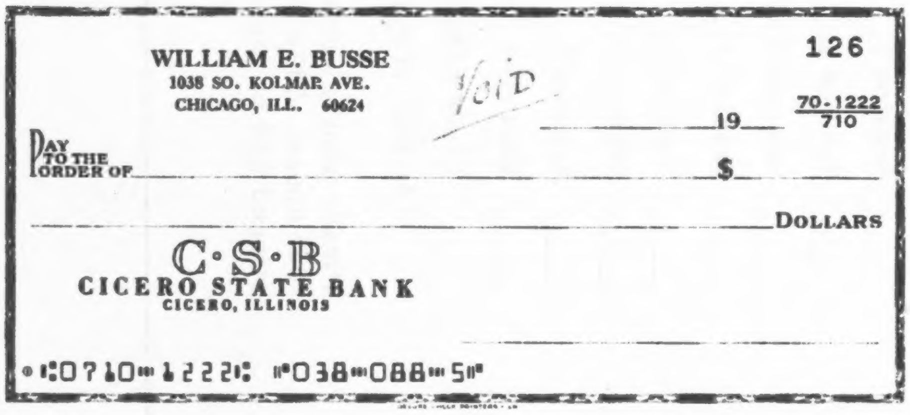
NANCY L. BARNSTABLE
1025 N. California Ave.

F/W, 15, 5-4, 100, slender build, olive comp brown eyes, dk brown hair. Wearing orange print skirt, bright green blouse, white sox, black shoes, green jacket. Having family difficulties and is roaming neighborhood of Old Town on Wells St. Missing since 25 April 67. RD F124807. Area 4 Youth Sec.

WANTED FOR THEFT INVESTIGATION:

John Terrell Stewart, LKA: 1637 N. Talman
M/W, 16, 5-6, 140, medium build, medium comp, green eyes, brn hair, clean shaven, speaks with southern accent. Wore Brown Levi type trousers & green tee shirt. Wanted for investigation of theft committed at 3204 W. Carmen Ave. R.D. F181313 – Gen. Assgn. Section, Area 5.

TAKEN IN A BURGLARY



WILLIAM E. BUSSE
1038 SO. KOLMAR AVE.
CHICAGO, ILL. 60624

126

Void

_____ 19___ 70-1222/710

PAY TO THE ORDER OF _____ $_____

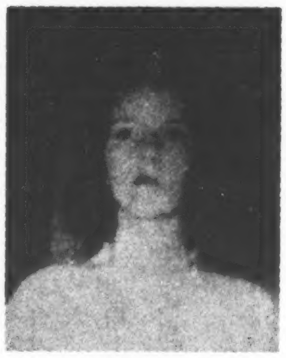
_____ DOLLARS

C·S·B
CICERO STATE BANK
CICERO, ILLINOIS

⑆0710⑉1222⑆ ⑈038⑉088⑉5⑈

22 checks, Nos 104 thru 125 inclusive. Drawn on Cicero State Bank, account of William E. Busse, 1038 N. Kolmar Ave. Taken in burglary of Kolmar Ave. premises betw. June 9th & 12th. Also taken, one Remington Elec Typewriter, Ser 3073224 brn & gray color adding machine. RD F187-209. Area 4 Burglary.

MISSING PERSON



CYNTHIA STAPLETON
5322 N. Winthrop Ave.

F/W, 16, 5-2, 110, slender build, fair comp, blue eyes, tinted red-brown hair, 2 front teeth dark, need capping. Wearing dark blue jeans, dk blue tennis shoes, color of blouse unknown. May be with one Edward Sabastian M/W, 19, in area of Lawrence & Broadway. Missing since 13 June 67. RD F189-416. Area 6 Youth Div.

DEATH NOTICES

John Notini, father of Ptlmn. Joseph Notini, TA4 and Ptlmn. Robert Notini, 20th Dist. passed away 24 Jun 67. Waked at Malec Funeral Home 834 N. Ashland after 1200 hrs 25 Jun. Funeral at Notre Dame Church, 1335 W. Harrison St. at 1000 hrs. Tues. 27 Jun. Interment Mt. Carmel Cemetery.

Mrs. Marjorie Wilson, mother of Det. Otis Wilson, DDA 1 Burglary, passed away Wed. 21 Jun 67 in Albion Mich. Waked at Bethel A.M.E. Church, 45th & Michigan, from 1900 to 2000 hrs only. Funeral Mon. 26 Jun at 1100 hrs from chapel. Interment Lincoln Cemetery. St. Jude League to meet Sun. 25 Jun at 1930 hrs.

Patrick Kelly, Ret. Ptlmn., old 36th Dist. passed away Thurs. 22 June 67. Waked at Cooney Funeral Home, 3552 N. Southport. Funeral mass 1000 hrs Mon 26 June 67 at St. Ann's Home, Techny, Ill. Interment All Saints Cemetery.

Timothy Breen, brother of Pat Breen, Dispatcher, Comm. Center, passed away 23 Jun 67 in London, England. Services and Interment 26 Jun 67 in London, England.

PATROLMAN CONFIRMATION CARD

(Save this card)

Name ______________________ Code No. __________

You have been selected from among the volunteers in your district for participation in the U. S. Department of Justice study of patrolmen, being conducted by the Industrial Relations Center, The University of Chicago. You have been scheduled to attend the indicated session, in accordance with your preferences.

_____	Session 1,	Mon, 20 March 1967 - 0900 hrs	Law School, Univ. of Chicago*
_____	Session 2,	Mon, 20 March 1967 - 1830 hrs	" "
_____	Session 3,	Wed, 22 March 1967 - 0900 hrs	" "
_____	Session 4,	Wed, 22 March 1967 - 1830 hrs	" "
_____	Session 5,	Thur, 23 March 1967 - 0900 hrs	Hall of Science, DePaul Univ. **
_____	Session 6,	Thur, 23 March 1967 - 1830 hrs	" "
_____	Session 7,	Mon, 27 March 1967 - 0900 hrs	" "
_____	Session 8,	Mon, 27 March 1967 - 1830 hrs	" "

* Law School, University of Chicago, 1111 East 60th Street, Classroom 4

** Hall of Science, DePaul University, 1036 West Belden Avenue, Room 300

Please plan to arrive promptly. The testing session will require approximately four hours. This card will serve as your admission ticket. Payment of volunteer's stipend will be made immediately after the completion of the testing session.

CHAPTER III--APPENDIX

DEPARTMENT SPECIAL ORDER	DATE OF ISSUE 4 February 1966	EFFECTIVE DATE 4 February 1966	NO. 66-7
SUBJECT PERFORMANCE RATINGS	DISTRIBUTION A	AMENDS	
REFERENCE	RESCINDS		

I. PURPOSE

This order provides for rating the performance of all members of the Department for the period 1 July to 31 December 1965.

II. AUTHORITY AND PROCEDURE

A. By authority of the Civil Service Commission, each department must send performance ratings to the Commission offices twice a year. One report covers the period from 1 January through 30 June; the second report covers the period from 1 July through 31 December.

B. Salmon colored Performance Rating Cards (CPD-14.270) have been prepared for the rating period 1 July to 31 December 1965.

C. The Director of Personnel will forward to each unit commander prepared rating cards for the members assigned as of 2 December 1965, a control sheet in duplicate with members' names, and a supply of blank cards.

D. The unit commander will use the control sheet as a work sheet for grading and as an aid in returning all cards to the Director of Personnel.

E. The unit commander will return all Performance Rating Cards and one copy of the control sheet, with the grades inserted in the "grade" column for each member, to the Special Projects and Research Section of the Personnel Division by 15 March 1966.

III. RESPONSIBILITIES OF COMMAND PERSONNEL

A. Command personnel are responsible for the administration of the Performance Rating System within units under their supervision. They will assure that the system is administered fairly; they will assure that large numbers of men are not set back by a policy of assigning unusually high performance marks to all members of a unit; they will assire that men deserving of a high grade are not given a low grade because a supervisor does not understand how the system operates.

B. Command personnel will use the computed averages to assure that the averages for the unit under their supervision are equitable. For example: watch commander "A" may rate his men and come up with an average of 85.00. Watch commander "B" may rate his men and come up with an average of 73.00. The district commander must be conscious of this variation. If the group under watch commander "A" is obviously superior in its performance to the group under watch commander "B", the district commander will be aware of this and accept the difference. If, on the other hand, there is no apparent difference in the two groups, it is the district commander's responsibility to bring about a more equitable basis of rating between the two watch commanders. In a similar manner, commanding officers at higher levels will be responsible for reviewing rating averages under their command to assure fairness.

IV. RESPONSIBILITIES OF UNIT AND WATCH COMMANDERS

A. A commander may appoint teams of two or three supervisors to rate their subordinates. Each team should operate under the direction of a higher level supervisor.

B. The captains and lieutenants in a unit are expected to confer to rate their sergeants. In this manner, the rating of the sergeants, as one is compared with the other, can be accomplished more equitably than if the rating of each sergeant was left up to his lieutenant. The same principle applies to higher ranks.

C. Commanders should be aware that the entire range of grades can be used. It can reasonably be expected that the average grades of each of the three bureaus will fall into the range of "good."

D. Patrol districts will treat each watch (1, 2, 3) as a separate unit and will enter the watch number in the blank area above the "final grade" box on the rating card. The watch commander will sign the cards as unit commander.

V. RESPONSIBILITIES OF THE RATER

A. Every rater will study the instructions in the Performance Rating Guide which is issued by the Civil Service Commission before making the performance evaluation.

B. A member will be rated by the supervisors of the unit to which the card is sent if the member was assigned there for a period of at least thirty calendar days during the rating period. When a member has worked for more than one supervisor during the rating period indicated, the previous supervisor will be contacted and his opinion of the member's performance will be considered in the final grade. Any personnel in the unit who were either on leave of absence or resigned from the Department during the period 1 July to 31 December 1965 will be rated if they were in the unit for at least thirty days. The unit commander will notify the Personnel Division on Pax 347 if a card is received for a member who cannot be rated according to this policy.

C. It is pertinent that the rater judge each member under his command according to the same set of fair and impartial standards.

D. The rater must consider the favorable and unfavorable repetitive characteristics and work habits of the members being rated. Upon indicating a member's grade for one of the five traits, the rater must determine whether the member has consistently warranted that grade (and corresponding appraisal term), or if the member has met the job requirements only occasionally during the rating period.

E. With reference to the trait of "Attendance and Promptness", the rater should be aware of the fact that a member taking sick leave is to be considered absent and such practice can considerably lower the performance rating grade. Excessive sick leave should not only affect the member's grade for "Attendance and Promptness", but it should be reflected in his grades for the traits of "Dependability" and "Quantity of Work. " The rater will consider grades in the "Fair or Unsatisfactory" category when the member's sick leave is excessive or there is an indication of sick leave abuse. (Refer to the memorandum pertaining to sick leave abuse that was sent to all units by the Director of Personnel on 3 May 1965.)

F. In units where several supervisors have supervised the same member, they will confer with each other to rate the member. In this manner, the ratings of the member, as he is compared with the others, can be accomplished more equitably than if the rating of each member was left up to one supervisor.

G. The immediate supervisor will enter the grades on the performance rating card in ink and will sign his name in the space designated on the card. The rating will be reviewed and signed in the appropriate space by the supervisor at the next higher level and the unit commander. It will be the duty of the unit commander to insure that the rating has been given careful consideration and is properly justified.

H. If approval is not indicated by the reviewers' signatures, a new performance card will be prepared with the amended grade. Both cards will be forwarded to the Personnel Division.

Authenticated by:

O. W. Wilson
Superintendent of Police

CHAPTER III--APPENDIX

PERFORMANCE RATING CHECKLIST/CHICAGO POLICE

EMPLOYEE'S NAME

UNIT

GRADE TITLE

RATING PERIOD FROM TO

SUPERVISOR

INSTRUCTIONS: This checklist guide has been constructed with a view toward assisting the rater in determining accurate and equitable numerical values for the five major traits on the Performance Rating Card and the resultant Final Grade.

Below are listed a number of items which could be considered in determining a grade for each of the five traits that comprise the Final Grade on the Performance Rating Card. Although there may be as many as 20 items for each trait you may select only those that pertain to a ratee's job classification in order to determine an average or a final numerical grade for each trait. You are asked to examine the items and select those that best evaluate the type of duties and responsibilities that the person to be rated performs. After the items have been selected, the rater should examine the rating values listed at the top of the page. If it is felt that the value of the item being examined falls into a certain range - - for instance, EXCELLENT (85 – 89) - - a definite numerical value should be established and indicated in the column next to the item rated. After each of the items under a trait have been graded, the values should be added and then divided by the number of items selected in order to determine an average. This would be the value placed on the Performance Rating Card for the respective major trait (Quality of Work, etc.) You then add the grades for the five traits on the Performance Rating Card, as done in the past, and determine an average mark which would be the Final Grade.

An additional number of items for supervisory personnel (preceded by an asterisk) are included below those items to be graded for non-supervisory employees; however, all items in the list should be considered if conducive to evaluating supervisory personnel.

In the rating of non-supervisory personnel those items preceded by an asterisk should be disregarded.

These checklist guides may be retained in your unit as a record for determining performance rating grades. They are not to be returned with the cards.

	UN-SATISFACTORY	FAIR	SATISFACTORY	GOOD	EXCELLENT	OUTSTANDING	SUPERIOR
	BELOW 70	70 74	75 79	80 84	85 89	90 94	95 +
QUALITY OF WORK							
ACCURACY							
JUDGMENT							
KNOWLEDGE OF DUTIES							
SYSTEMATIC							
THOROUGHNESS							
ABILITY TO LEARN							
INTEREST							
WILLINGNESS							
INITIATIVE							
EXTRACURRICULAR STUDY							
REPORTS NEAT, GRAMMATICALLY CORRECT							
COMPLIANCE WITH DEPARTMENT POLICY AND DIRECTIVES							
FELONY CONVICTION RECORD							
MISDEMEANOR CONVICTION RECORD							
TRAFFIC CONVICTION RECORD							
EMPHASIS ON ALL AREAS OF ENFORCEMENT							
KNOWS AREA TO WHICH ASSIGNED							
ATTENTION TO PRIMARY FUNCTIONS OF POSITION							
ENGAGES IN AGGRESSIVE, PREVENTIVE PATROL							
* OBTAINS DESIRED WORK RESULTS FROM SUBORDINATES							
* EXPLANATION OF WORK TO BE DONE							
* ALERT TO POSSIBLE PROBLEMS							
* SYSTEMATICALLY CHECKS SUBORDINATES WORK							
* ABILITY TO EXPLAIN THE REASONS FOR PROCEDURES OR ASSIGNMENTS							
* KNOWLEDGE OF, AND CONFORMITY TO DEPARTMENTAL DIRECTIVES							
* READILY ASSUMES RESPONSIBILITY FOR DISCIPLINARY ACTION WHEN NEED ARISES							
* ABILITY TO IMPROVISE							
TOTAL							

GRAND TOTAL	DIVIDED BY NO. OF ITEMS USED	FINAL GRADE

	UN-SATISFACTORY	FAIR	SATISFACTORY	GOOD	EXCELLENT	OUTSTANDING	SUPERIOR
	BELOW 70	70 74	75 79	80 84	85 89	90 94	95 +
QUANTITY OF WORK							
PRODUCTION							
ORGANIZATION OF TIME							
INDUSTRIOUSNESS							
ASSISTANCE, VOLUNTARY AND WHEN REQUESTED							
EFFORTS BEYOND BASIC REQUIREMENTS							
ACCOMPLISHMENTS							
FELONY ARREST RECORD							
MISDEMEANOR ARREST RECORD							
TRAFFIC ENFORCEMENT RECORD							
CURFEW VIOLATION ARRESTS							
* AMOUNT OF WORK PRODUCED BY UNIT							
* COORDINATES ACTIVITIES OF SUBORDINATES							
* CHECKS ADEQUACY AND PROGRESS OF WORK PRODUCED BY SUBORDINATES							
* ARRANGES TIME OF SUBORDINATES SO AS TO AVOID IDLENESS							

(TOTAL CONTINUED ON REVERSE SIDE)

CPD-11.259 (6/66)

* To be considered with preceeding items in evaluating supervisory personnel.

	UN-SATISFACTORY BELOW 70	FAIR 70 74	SATISFACTORY 75 79	GOOD 80 84	EXCELLENT 85 89	OUTSTANDING 90 94	SUPERIOR 95 +
QUANTITY OF WORK (CONTINUED)							
TOTAL							

GRAND TOTAL	DIVIDED BY NO. OF ITEMS USED	FINAL GRADE

DEPENDABILITY

RELIABLE							
REQUIRES LITTLE SUPERVISION							
FOLLOWS INSTRUCTIONS OR ORDERS							
ADAPTS TO CHANGE							
CONSISTENT IN PERFORMANCE							
ACCEPTS RESPONSIBILITY							
COOPERATIVE							
CARE OF DEPARTMENT EQUIPMENT							
INTEGRITY							
DISCIPLINARY RECORD							
ATTITUDE TOWARD DEPARTMENT GOALS AND OBJECTIVES							
ADJUSTED PERSONAL LIFE							
ACCEPTS AND ACTS ON CONSTRUCTIVE CRITICISMS OF SUPERIORS							
* WILLING TO ASSUME RESPONSIBILITY							
* DELEGATES AUTHORITY WHEN POSSIBLE							
* WILLINGNESS TO TAKE CORRECTIVE MEASURES WITH LAX OR DEFICIENT SUBORDINATES							
* INSURES THAT SUBORDINATES PROPERLY CARE FOR MATERIALS AND EQUIPMENT							
TOTAL							

GRAND TOTAL	DIVIDED BY NO. OF ITEMS USED	FINAL GRADE

PERSONAL RELATIONSHIPS

TACTFUL							
TEAM MEMBER							
EVEN-TEMPERED							
SELF-CONTROL							
PATIENT							
REACTION TO STRESS							
DEMEANOR							
SELF-RESPECT							

	UN-SATISFACTORY BELOW 70	FAIR 70 74	SATISFACTORY 75 79	GOOD 80 84	EXCELLENT 85 89	OUTSTANDING 90 94	SUPERIOR 95 +
ABLE TO GAIN OTHERS' RESPECT							
RESPECTS RIGHTS OF OTHERS							
MAKES FAVORABLE IMPRESSIONS							
ACCEPTANCE OF CORRECTION AND GUIDANCE							
UNIFORM OR DRESS							
PROPER APPEARANCE							
COURTESY							
CONSIDERATION							
OBJECTIVITY							
RESPECT FOR PROPERTY OF OTHERS							
ABILITY TO DEAL WITH THE PUBLIC							
* ABILITY TO GAIN SUBORDINATES' RESPECT							
* IMPARTIALITY							
* INTERESTED IN PROGRESS OF SUBORDINATES							
* MAINTAINS HIS PERSONAL APPEARANCE SO AS TO SERVE AS AN EXAMPLE FOR SUBORDINATES							
* MAKES FACE-TO-FACE CORRECTIONS WITH EQUANIMITY AND OBJECTIVITY							
TOTAL							

GRAND TOTAL	DIVIDED BY NO. OF ITEMS USED	FINAL GRADE

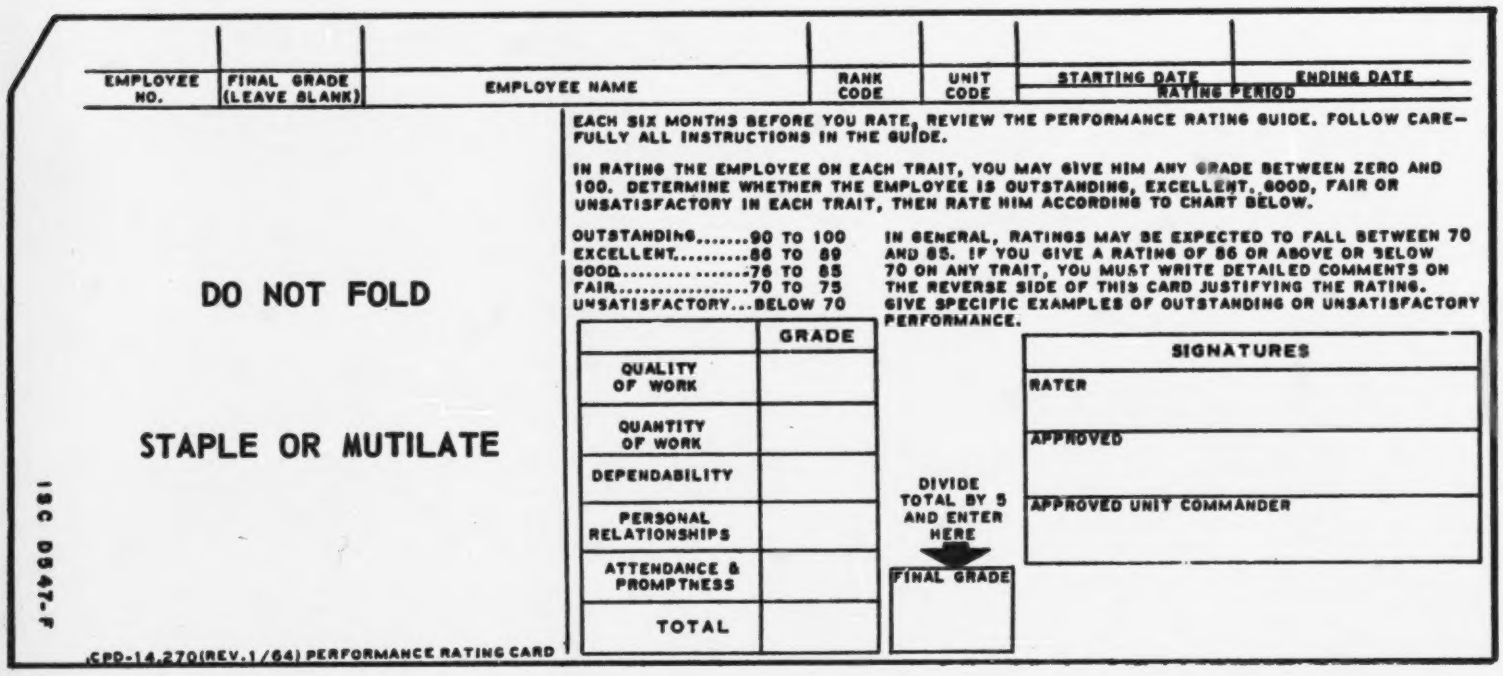
ATTENDANCE AND PROMPTNESS

REPORTS FOR WORK ON TIME							
DOES NOT LEAVE EARLY							
ABSENTEE RECORD							
TAKES ONLY ALLOTTED LUNCH AND REST TIME							
DOES NOT USE DUTY TIME FOR PERSONAL CHORES OR ERRANDS							
HANDLES OR RESPONDS TO ASSIGNMENTS PROMPTLY							
SUBMITS REPORTS PROMPTLY							
PHYSICAL CONDITION WARRANTS AVAILABILITY FOR ANY ASSIGNMENTS							
* SETS A GOOD EXAMPLE FOR SUBORDINATES IN THIS AREA							
* EXACTS GOOD ATTENDANCE AND PROMPTNESS FROM UNIT MEMBERS							
* IS READILY AVAILABLE WHEN NEEDED- MAKES NO EFFORT TO AVOID UNPLEASANT TASKS							
* UNIT COMPLETES ASSIGNMENTS WITHOUT DELAY							
TOTAL							

GRAND TOTAL	DIVIDED BY NO. OF ITEMS USED	FINAL GRADE

* To be considered with preceeding items in evaluating supervisory personnel.

2



EMPLOYEE NO.	FINAL GRADE (LEAVE BLANK)	EMPLOYEE NAME	RANK CODE	UNIT CODE	STARTING DATE	ENDING DATE
					RATING PERIOD	

DO NOT FOLD

STAPLE OR MUTILATE

EACH SIX MONTHS BEFORE YOU RATE, REVIEW THE PERFORMANCE RATING GUIDE. FOLLOW CAREFULLY ALL INSTRUCTIONS IN THE GUIDE.

IN RATING THE EMPLOYEE ON EACH TRAIT, YOU MAY GIVE HIM ANY GRADE BETWEEN ZERO AND 100. DETERMINE WHETHER THE EMPLOYEE IS OUTSTANDING, EXCELLENT, GOOD, FAIR OR UNSATISFACTORY IN EACH TRAIT, THEN RATE HIM ACCORDING TO CHART BELOW.

OUTSTANDING	90 TO 100
EXCELLENT	86 TO 89
GOOD	76 TO 85
FAIR	70 TO 75
UNSATISFACTORY	BELOW 70

IN GENERAL, RATINGS MAY BE EXPECTED TO FALL BETWEEN 70 AND 85. IF YOU GIVE A RATING OF 86 OR ABOVE OR BELOW 70 ON ANY TRAIT, YOU MUST WRITE DETAILED COMMENTS ON THE REVERSE SIDE OF THIS CARD JUSTIFYING THE RATING. GIVE SPECIFIC EXAMPLES OF OUTSTANDING OR UNSATISFACTORY PERFORMANCE.

	GRADE
QUALITY OF WORK	
QUANTITY OF WORK	
DEPENDABILITY	
PERSONAL RELATIONSHIPS	
ATTENDANCE & PROMPTNESS	
TOTAL	

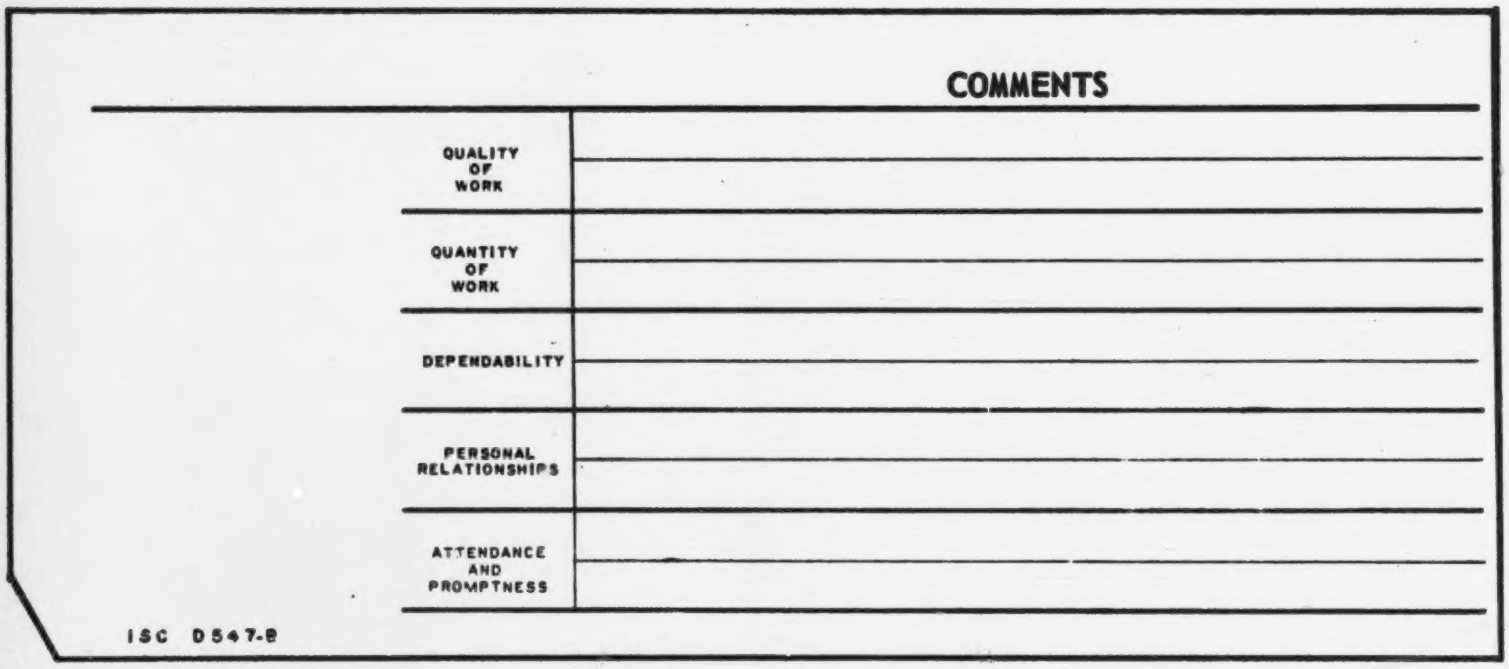
DIVIDE TOTAL BY 5 AND ENTER HERE

FINAL GRADE

SIGNATURES
RATER
APPROVED
APPROVED UNIT COMMANDER

ISC D547-F

CPD-14.270(REV.1/64) PERFORMANCE RATING CARD



COMMENTS

QUALITY OF WORK	
QUANTITY OF WORK	
DEPENDABILITY	
PERSONAL RELATIONSHIPS	
ATTENDANCE AND PROMPTNESS	

ISC D547-B

CHAPTER IV

Selection and Description of the Psychological Test Batteries

SELECTION OF THE TESTS

One of the essential requirements for attaining the first objective of the study--the prediction of patrolman performance--was the identification and utilization of psychological tests judged to measure the personal skills and attributes which have a direct bearing on the success of the patrolman in the performance of his duties. The selection of the tests was also crucial for the second objective of the study--the identification of specific patterns of good and poor patrolman performance. In order to use these patterns for practical purposes of selection, placement, duty assignment, and promotion, definitive and differentiating psychological descriptions of the patrolmen who exhibit them are needed.

The occupational analysis discussed in Chapter II revealed many of the critical behavioral requirements of the police officer's job. Also our field experience with Chicago police officers provided significant knowledge of the complex demands which confront the patrolman. Our resulting insights supplied a foundation for evaluating the relative merits of each particular test and its appropriateness for inclusion in the test battery. Another source of information for selection tests was the collective experience of members of the Employee Appraisal Section of the Chicago Police Department Personnel Division. These psychologists had extensive experience in the assessment of police officers, which aided considerably in the decision-making process. A third source of information was the staff of the Personnel Division itself, specifically the Director of the Personnel Division and his Assistant.

Thus, the staffs of the Measurement Research Division and the Personnel Division collaborated on the composition of the test battery, with primary responsibility resting with Measurement Research Personnel. The strategy was to consider all possible tests, using the Sixth Mental Measurements Yearbook (Buros, 1965) and current catalogs of test publishers as sources. Particular emphasis was placed on tests developed by the Industrial Relations Center and validated in industrial organizations during the past two decades, tests currently in use by the Employee Appraisal Section, and tests being used by other agencies and organizations affiliated with the Department (e.g., Chicago Civil Service Commission and The McMurry Company). In addition to the essential requirement of relevance, a number of other technical and practical considerations were taken into account in selecting the tests. In view of the large number of patrolmen to be tested in the study--and the even larger number who would be involved if the validated battery were used for the selection of recruits by the Chicago Police Department or other police departments in the country--the following criteria were applied in choosing the tests:

1. Group vs. Individual Administration. Only tests which could be administered on a group basis were included in the battery. This excluded such tests as the Wechsler Adult Intelligence Scale and many of the projective tests, which must be administered to individuals by highly trained and specialized psychologists.
2. Paper-and-Pencil vs. Apparatus Tests. Paper-and-pencil tests were used in preference to tests requiring special apparatus for their administration, such as the peg board or pursuit rotor tests.
3. Objective vs. Subjective Scoring. Tests were selected with a view to the objectivity of their scoring procedures. Thus, tests which could be directly scored with a prepared key were used rather than tests such as the Rorschach, which are most effective when the obtained indices are subjectively interpreted.

In general, the tests are practical and fairly economical with respect both to the subjects' time and to the time and expense incurred by the organization for administration and scoring. All of them can be administered and scored by clerical personnel trained for this purpose. The interpretation of test results, of course, should always be made by professional personnel.

No test construction was attempted for this study. However, it is probable that battery validity could be improved by the construction of specialized tests for police-officer selection. It was thought that one of the important contributions of this study would be to identify those areas which promised to be most fruitful for the construction of new tests.

Chicago's police officers may differ from those in other municipalities in that they have passed through an intensive screening procedure (Appendix) before being appointed to the force. Because of this pre-screening, it was decided to exclude tests specifically aimed at assessment of a global intelligence quotient (since the Civil Service Commission essentially screened out individuals with an I.Q. of less than 100) and at direct assessment of emotional health (since the Department itself utilizes psychiatric screening). We felt that these screening procedures were sound and that additional testing in these areas would add little to our analysis, since those obviously unfit on the score of I.Q. or adjustment had already been screened out. If the population had not been pre-screened in these areas, we probably could have improved our validities by including additional mental ability and adjustment measures.

It would have been ideal, of course, to have been able to include tests which had demonstrated some validity for the prediction of patrolman performance in other studies. Unfortunately, positive evidence of this sort was largely lacking. However, in view of the exploratory nature of this study and keeping in mind the requirements of relevance, as indicated by the occupational analysis, and the practical considerations previously discussed, we decided we would cover as wide a span of human behavior as possible. The tests chosen can be classified into the following behavioral areas:

1. Motivational Measures
   - --Dimensions of objective background data.
   - --Dimensions of work interests indicating strength, flexibility, and vocational aspiration level of occupational interests.
2. Intellectual Measures
   - --Primary mental abilities in the areas of reasoning, language facility, and visual perception.
   - --Special aptitudes, such as creative potential and insight in social situations.
3. Behavioral Measures
   - --Dimensions of the relatively permanent temperament traits or characteristic modes of response of individuals in both normal and pressure situations.
   - --Dimensions of personality functioning as measured in tests purporting to represent various personality systems.

The tests used in the selection batteries for the Wave I and Wave II samples of patrolmen, together with the number of dimensions or variables included in each test and the administration time for the batteries, are shown in Table 1. Short descriptions of each of the tests used, including information about the authors, publishers, definitions of the dimensions measured, time limits, and sample items, are given on the remaining pages of this chapter. The Appendix contains a description of the test administration procedure, including the orientation statement given to the participants.

CHAPTER IV -- Table 1

Test (Predictor) Variables for Wave I and Wave II Samples of Patrolmen

AREA OF MEASUREMENT	TEST USED	NUMBER OF VARIABLES	WAVE I SAMPLE	WAVE II SAMPLE	COMMON TESTS
MOTIVATION					
Background & Experience	PHI	15	XX	XX	XX
Work Interest	WII	14		XX	
INTELLECTUAL SKILL					
Reasoning	NVR	1	XX		
Language	UC	1	XX		
Perception	CS	1	XX	XX	XX
	CF	1	XX	XX	XX
	PS	1	XX	XX	XX
APTITUDE					
Creative	AC	2	XX		
	CREE	14		XX	
Social	SI	8	XX	XX	XX
BEHAVIOR					
Temperament	TC	24	XX	XX	XX
	EPPS	15		XX	
Pressure Tolerance	PRESS	5	XX	XX	XX
Personality Systems	BI	1	XX		
	HAND	8	XX		
	MPI	3	XX		
	A-D	7	XX	XX	XX
TOTAL NUMBER OF TESTS			14	11	8
ADMINISTRATION TIME (HOURS)			4	4	3

DESCRIPTION OF THE TESTS

TESTS OF MOTIVATION

PERSONAL HISTORY INDEX
[PHI]

AUTHORS: Melany E. Baehr, Ph.D., Robert K. Burns, Ph.D., and Robert N. McMurry, Ph.D., Industrial Relations Center, The University of Chicago.

MEASURES: This Index assists in predicting future job success on the basis of past performance and experience. Past performance and experience are evaluated by an analysis of the scores obtained on 15 performance factors. These performance factors are derived from 87 factual items of information concerning family, education, and work history; they were identified through a series of factorial studies. Their names and interpretations are given below:

School Achievement
- --academic achievement
- --a liking for, and adjustment to, the school environment

Higher Educational Achievement
- --higher-level academic specialization
- --late vocational start and marriage

Drive
- --motivation to achieve success
- --preference for supervising others

Leadership & Group Participation
- --desire for interpersonal contact
- --interest in influencing others

Financial Responsibility
- --satisfactory handling of personal budget
- --motivation for financial achievement

MEASURES:
(cont.)

Early Family Responsibility
--early marriage and establishment of a family
--interest in family activities

Parental Family Adjustment
--development of positive attitudes in childhood
--satisfactory relationships in parental family

Stability
--established stability through successful work history
--concern for maintenance of status quo

School Activities
--academic achievement
--participation in athletic and social activities

Professional-Successful Parents
--self-employed or professional successful father
--happy, comfortable home life

Educational-Vocational Consistency
--preferred occupations closely related and in line with education

Vocational Decisiveness
--decisive choice of an occupation
--early qualification and start

Vocational Satisfaction
--satisfaction with occupation
--peak performance seen as in the future

Selling Experience
--experience in sales, including door-to-door selling and real estate transactions

General Health
--long-term, better than average health
--little lost work time because of illness

SAMPLE ITEMS:

---

34. Do you:
a. ____Own a house
b. ____Rent a house or apartment
c. ____Rent a room
d. ____Live with relatives
e. ____Other

SAMPLE ITEMS:
(cont.)

---

65. While in high school, where did you rank in scholarship in your class?

a. ____Upper 10%
b. ____Upper 25%
c. ____Upper 50%
d. ____Lower 50%
e. ____Don't know

TIME LIMITS: Untimed. Approximately 10 to 20 minutes.

WORK INTEREST INDEX (Non-Verbal)
[WII]

AUTHORS: Melany E. Baehr, Ph.D., Richard Renck, Ph.D., and Robert K. Burns, Ph.D., Industrial Relations Center, The University of Chicago.

MEASURES: This Index is pictorial and completely non-verbal. It consists of 96 pictures, each of which shows a man engaged in a specific occupation. For each picture the subjects checks whether he would like "L" or dislike "D" the work shown. The principal purpose of the Index is to assist in determining an individual's pattern of occupational interests through an analysis of the scores he obtains on 12 work interest factors and two work attitude scales (Aspiration Level and Flexibility of Interests). The 12 factors were defined through a series of factorial studies. They are named and interpreted as follows:

Professional & Technical. Professional and technical skills used for therapeutic purposes--physical therapist, dentist.

Social & Verbal. Social and verbal skills used in teaching or counseling and resulting in the esteem of others--school principal, office manager.

Authority & Prestige. Authority and prestige or independent judgment used in unpredictable or non-routine situations--civil lawyer, newspaper reporter.

Artistic & Interpretative. Artistic and interpretative talents used in the performing arts--actor, composer.

MEASURES:
(cont.)

Artistic & Stylized. Artistic ability used to make stylized products--taxidermist, glassblower.

Artistic & Creative. Artistic or creative talent used in the communication of ideas--sculptor, commercial designer.

Technical & Scientific. Technical and scientific skills used in the physical or biological sciences--physicist, pharmacist.

Clerical & Routine. Routine clerical skills used in concrete procedures--general office worker, inventory clerk.

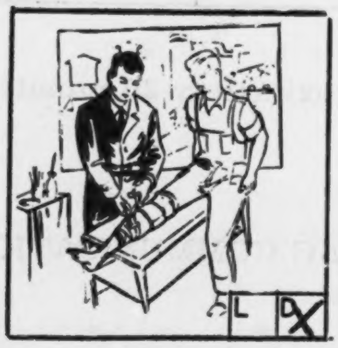
Business Contact & Structured. Business contact in structured social situations--hotel desk clerk, grocery checker.

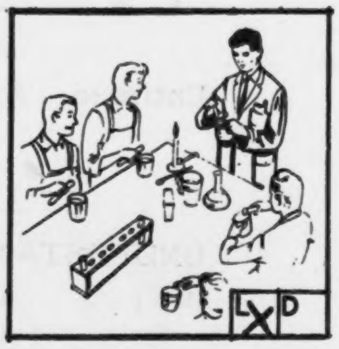
Personal Service & Persuasive. Personal service for the advantage or enjoyment of others--ship's purser, welfare director.

Mechanical & Productive. Mechanical skills for producing or repairing a product--automobile mechanic, cabinet maker.

Control of Massive Equipment. Direction of massive equipment in largely masculine settings--locomotive engineer, cable driller.

SAMPLE ITEMS:





TIME LIMITS: Untimed. Approximately 10 to 15 minutes.

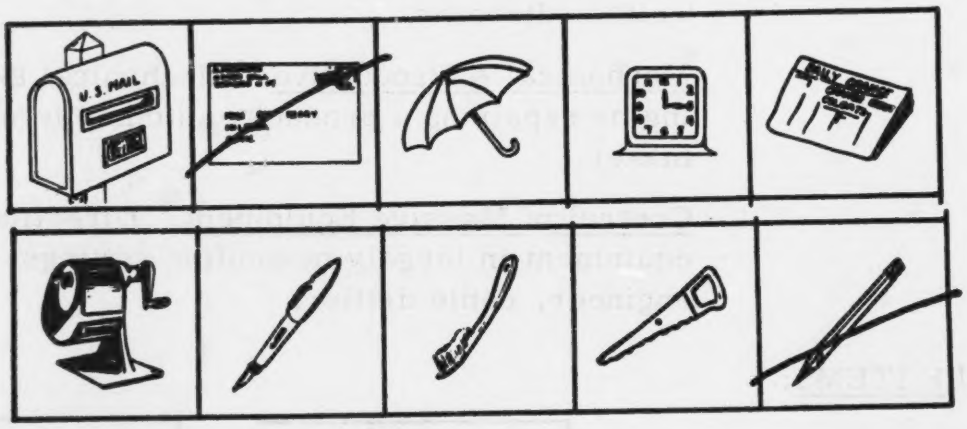
TESTS OF INTELLECTUAL SKILL

NON-VERBAL REASONING
[NVR]

AUTHOR: Raymond J. Corsini, Ph. D., Industrial Relations Center, The University of Chicago.

MEASURES: This test measures reasoning ability through the medium of pictorial problems. Each of the 44 test items consists of a row of five pictures. The subject is asked to find "the picture that goes best with the first picture in the row." He then draws a line through the picture he selects. The test yields a single score.

SAMPLE ITEMS:



TIME LIMITS: Untimed. Approximately 20 minutes.

UNDERSTANDING COMMUNICATION
[UC]

AUTHOR: T. G. Thurstone, Ph.D., The School of Education, The University of North Carolina. Published by Industrial Relations Center, The University of Chicago.

MEASURES: This test measures comprehension of verbal material in the form of short sentences and phrases. Vocabulary and speed of reading are of minor importance in achieving a good verbal comprehension score on it. Test items are incomplete sentences with four alternative completions. The subject selects the word or phrase which correctly completes the sentence.

SAMPLE ITEMS:

When waves hit against rocks, small pieces of the rock are broken off. When these broken pieces rub against other rocks, they break up again, making

A. rains. B. sand. C. waves. D. floods.

If the pilot of an airliner meets too much ice, he can ascend or descend to different flight levels until he finds a warmer

A. engine. B. transport. C. temperature. D. day.

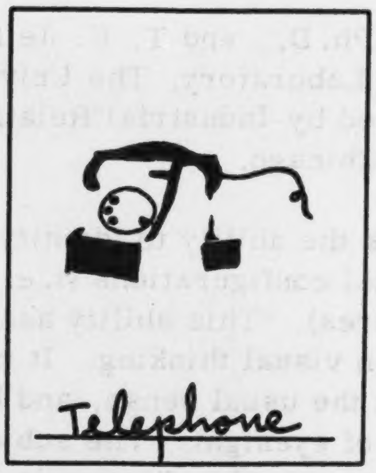
TIME LIMITS: 15 minutes.

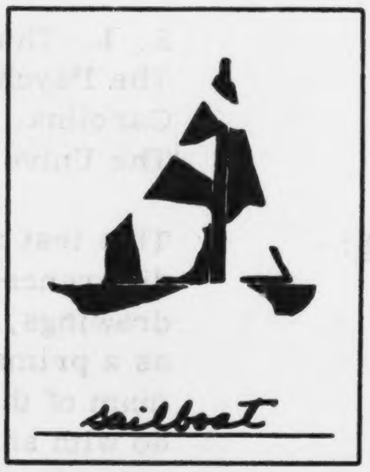
CLOSURE SPEED (Gestalt Completion)
[CS]

AUTHORS: L. L. Thurstone, Ph.D., and T. E. Jeffrey, Ph.D., The Psychometric Laboratory, The University of North Carolina. Published by Industrial Relations Center, The University of Chicago.

MEASURES: Closure speed is the ability to see an apparently disorganized or unrelated group of parts or ideas as a meaningful whole or single unit. In other words, it is the ability to unify a complex situation. In this test, the subject is asked to identify objects which are only partially or sketchily pictured.

SAMPLE ITEMS:





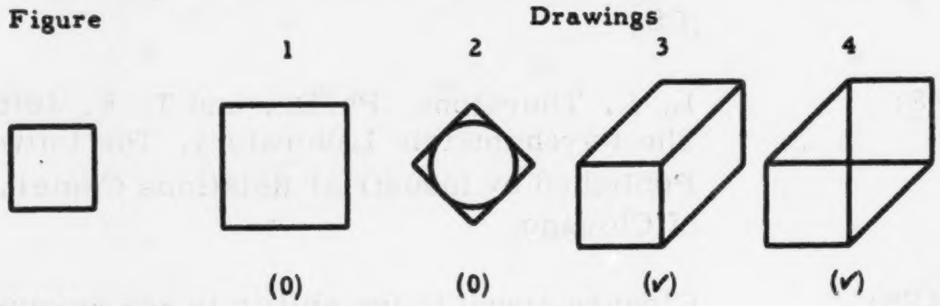
TIME LIMITS: 3 minutes.

CLOSURE FLEXIBILITY (Concealed Figures) [CF]

AUTHORS: L. L. Thurstone, Ph.D., and T. E. Jeffrey, Ph.D., The Psychometric Laboratory, The University of North Carolina. Published by Industrial Relations Center, The University of Chicago.

MEASURES: This is an objective test of the ability to keep a figure in mind in the face of distraction, that is, to see a given figure (diagram, drawing, or configuration of some kind) which is "hidden" or embedded in a larger, more complex, drawing, diagram, or configuration. The subject is given a figure followed by four drawings, and asked to indicate in which of the drawings the figure is "concealed." (The figure must retain its size and general spatial orientation.)

SAMPLE ITEMS:



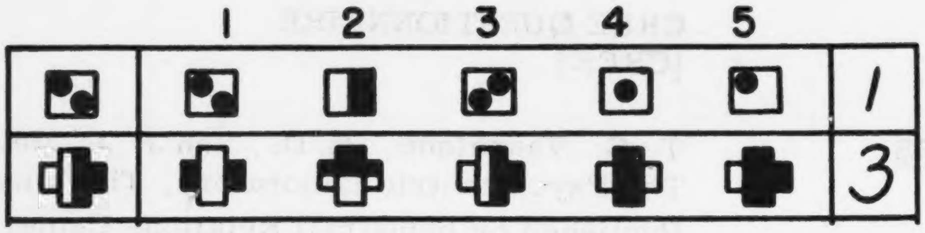
TIME LIMITS: 10 minutes.

PERCEPTUAL SPEED (Identical Forms) [PS]

AUTHORS: L. L. Thurstone, Ph.D., and T. E. Jeffrey, Ph.D., The Psychometric Laboratory, The University of North Carolina. Published by Industrial Relations Center, The University of Chicago.

MEASURES: This test measures the ability to identify likenesses and differences in visual configurations (i.e., diagrams, drawings, and figures). This ability has been identified as a primary one in visual thinking. It requires a minimum of thinking in the usual sense, and has very little to do with sharpness of eyesight. The subject is asked to indicate which of five similar figures is identical with a figure on the left. The test yields a single score.

SAMPLE ITEMS:



TIME LIMITS: 5 minutes.

TESTS OF APTITUDE

AC TEST OF CREATIVE ABILITY (Revised Short Form A) [AC]

AUTHOR: AC Spark Plug Division of General Motors Corporation. Published by Industrial Relations Center, The University of Chicago.

MEASURES: The AC Test is an objective test of creative potential. It assesses both the quantity and the uniqueness of the ideas an individual can produce in a given situation with given time limits. The total test has five parts and two alternate forms. There are also two alternate short forms containing only three parts. For the purposes of the Chicago Police Department study, only Parts II and III of Revised Short Form A were used. Instructions for these two parts are given below with a sample item from each.

SAMPLE ITEMS:

II. Below are listed a number of statements which you are to assume are true. Give as many reasons or explanations as you can to explain the truth of these statements.

April is the month when the fewest accidents of any kind occur in the U.S.

III. Below are listed five common objects. List all the possible uses to which these objects might be put (both uses that you have seen and uses that you can imagine).

A wooden ruler

TIME LIMITS:

Part II	10 minutes
Part III	15 minutes

CREE QUESTIONNAIRE
[CREE]

AUTHORS: T. G. Thurstone, Ph.D., and J. J. Mellinger, Ph.D., The Psychometric Laboratory, The University of North Carolina. Published by Industrial Relations Center, The University of Chicago.

MEASURES: This questionnaire was developed specifically to assess some of the significant aspects of temperament and personality which distinguish between creative and non-creative individuals. There are 145 questions, which the subject answers by circling "Y" (Yes), "N" (No), or "?" (Undecided). Results yield an overall creativity score plus scores on 13 creativity factors which are classified into four general areas of life functioning.

Social Orientation

Dominance vs. Submission--social dominance expressed through social leadership, a liking for communicating in formalized situations where the speaker is in control, and self-confidence.

Indifference vs. Involvement--asocial behavior, lack of social involvement, and nonparticipation in sports.

Independence vs. Conformity--independence, antisociability, a serious outlook, and a sense of being different.

Work Orientation

Unstructured vs. Structured Situation--preference for unsystematic, unstructured work situations where autonomous behavior is possible.

Unsystematic, Selective vs. Systematic, Prescribed Activity--individualistic approach to work, unsystematic work habits, and a low level of activity.

Personally Involved vs. Detached Attitude--deep involvement with job at hand and a disinclination for making quick decisions.

Pressure vs. Relaxed Situation--ability to produce ideas better while working under pressure.

MEASURES: (cont.)

Internal Functioning

High Energy vs. Low Energy Level--hyperactive, does things faster than most people, and usually in a hurry.

Fast vs. Slow Reaction--decisive, quick-thinking behavior.

High Ideational vs. Low Ideational Spontaneity--constantly beset by ideas in almost every situation.

Interests & Skills

High vs. Low Scientific & Theoretical

High vs. Low Artistic

High vs. Low Mechanical

SAMPLE ITEMS:

79. Do you like to work with theoretical ideas?. [✓ Y] ? N
101. Does working under pressure bother you?. . .Y [✓ ?] N
123. Do you enjoy promoting a new project?. . . . [✓ Y] ? N

TIME LIMITS: Untimed. Approximately 10 to 15 minutes.

TEST OF SOCIAL INSIGHT (Adult Edition)
[SI]

AUTHOR: Russell N. Cassel, Ed.D. Published by Martin M. Bruce, New Rochelle, New York.

MEASURES: This test appraises the characteristic mode of reaction an individual exhibits in dealing with interpersonal or social problems. It consists of 60 items, each of which describes an interpersonal problem and offers five alternative actions to take in handling it. The five alternatives cover the following five modes of resolving such problems:

Withdrawal--avoiding or escaping the problem by leaving the situation.

Passivity--remaining in the situation but taking no action to solve the problem.

Cooperation--initiating an active endeavor to solve the problem.

Competition--drawing the attention of others, attempting to deal with the problem by excelling or outdoing others.

MEASURES: (cont.)

Aggression--directing little effort to the problem but expressing hostility or strong feelings in response to it.

The frequency with which an individual selects each mode establishes his characteristic pattern in dealing with social and interpersonal problems.

SAMPLE ITEM:

After waiting in a very crowded restaurant for 20 minutes for the waitress to bring your order, she arrives with the wrong order. What do you do?

a. Insist that she make the correction immediately.
b. Walk out, leave the wrong order on table, and do not pay for it.
c. Say nothing and accept the order.
d. Report the error to the manager.
e. See if you can make satisfactory exchanges at your table.

TIME LIMITS: Untimed. Approximately 35 minutes.

TESTS OF BEHAVIOR

TEMPERAMENT COMPARATOR
[TC]

AUTHORS: Melany E. Baehr, Ph.D., and R. W. Pranis, Industrial Relations Center, The University of Chicago.

MEASURES: The Temperament Comparator measures the relatively permanent temperament traits of the individual. It yields measures of 18 temperament traits and of five temperament factors, all within the normal range of behavior.

MEASURES: (cont.)

The traits are:

Calm
Cautious
Decisive
Demonstrative
Emotionally Stable
Energetic
Enthusiastic
Even-Tempered
Lively
Persevering
Prompt Starter
Quick Worker
Seeks Company
Self-Confident
Serious
Socially at Ease
Steady Worker
Talkative

The five temperament factors are based on weighted combinations of different sets of these traits. The five factors are defined as follows:

I. Controlled (vs. Outgoing). Persons scoring high on this factor are cautious, serious, and undemonstrative and show steady application and perseverance in all they undertake. The opposite pole of this factor describes persons who are impulsive, happy-go-lucky, and demonstrative. They tend to do well in persuasive fields involving contact with other people and poorly in routine work or work requiring long periods of steady concentration.

II. Stable (vs. Unstable). Persons scoring high on this factor are calm, even-tempered, and of a generally cheerful disposition. They are cool and considered in their judgment and can be relied upon to keep their heads in trying conditions. Research has not clearly defined the opposite pole of this factor, although there is some indication of a behavior pattern characterized by nonadjustive emotional responses.

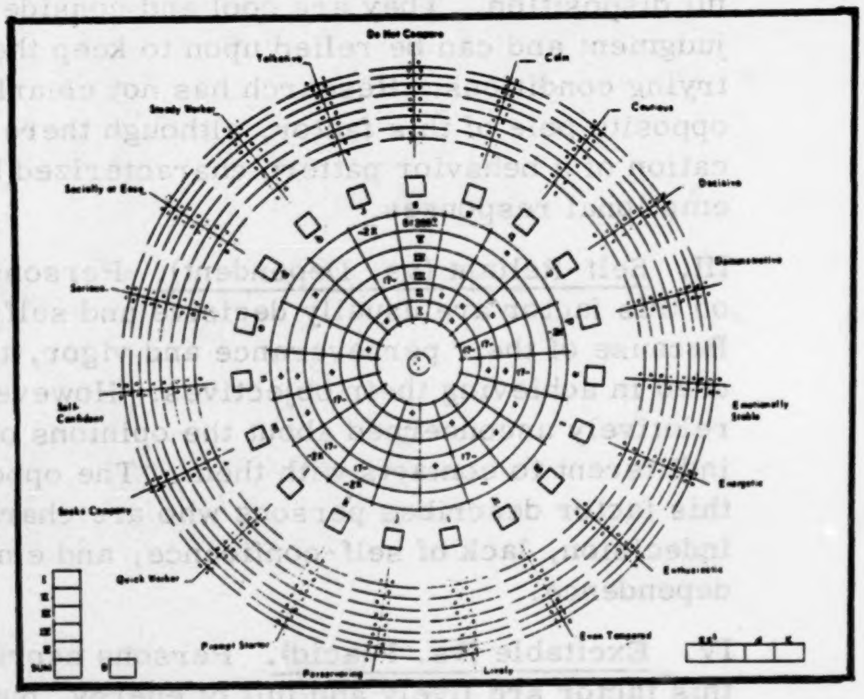
III. Self-Reliant (vs. Dependent). Persons scoring high on this factor are usually decisive and self-confident. Because of their perseverance and vigor, they often succeed in achieving their objectives. However, they are relatively unconcerned about the opinions of others and indifferent to contacts with them. The opposite pole of this factor describes persons who are characterized by indecision, lack of self-confidence, and emotional dependence.

IV. Excitable (vs. Placid). Persons scoring high on this factor are lively and full of energy, but this energy is splintered or dissipated through changeable and often erratic behavior. The opposite pole of the factor describes persons with a generally low level of physical activity and of emotional response.

V. Sociable (vs. Solitary). Persons scoring high on this factor seek social contacts but are slow to get down to the job on hand. The opposite pole of the factor describes persons who are often ill at ease in social situations but who are prompt starters and quick workers in the right environment. They may adjust well to work that is "solitary" and does not involve many social contacts.

The items in the Comparator consist of all possible pairings of the 18 temperament traits. The subject rotates a disc to successive positions within a circle which has the traits printed on its circumference. Bars on the disc link pairs of traits. The subject examines each pair of traits and places a mark in a hole on the disc near the trait which he has selected as more descriptive of his behavior. The Comparator also yields a Coefficient of Consistency which indicates the extent to which the subject was consistent in his selections or judgments.

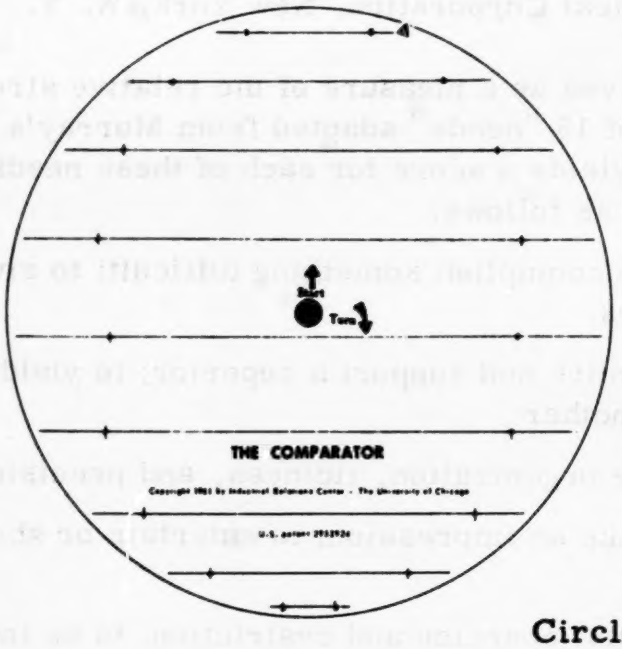
SAMPLE ITEMS:


Do Not Compare
Talkative
Calm
Steady Worker
Cautious
Socially at Ease
Decisive
Demonstrative
Serious
Self-Confident
Emotionally Stable
Energetic
Seeks Company
Enthusiastic
Quick Worker
Even Tempered
Prompt Starter
Persevering
Lively
SUM

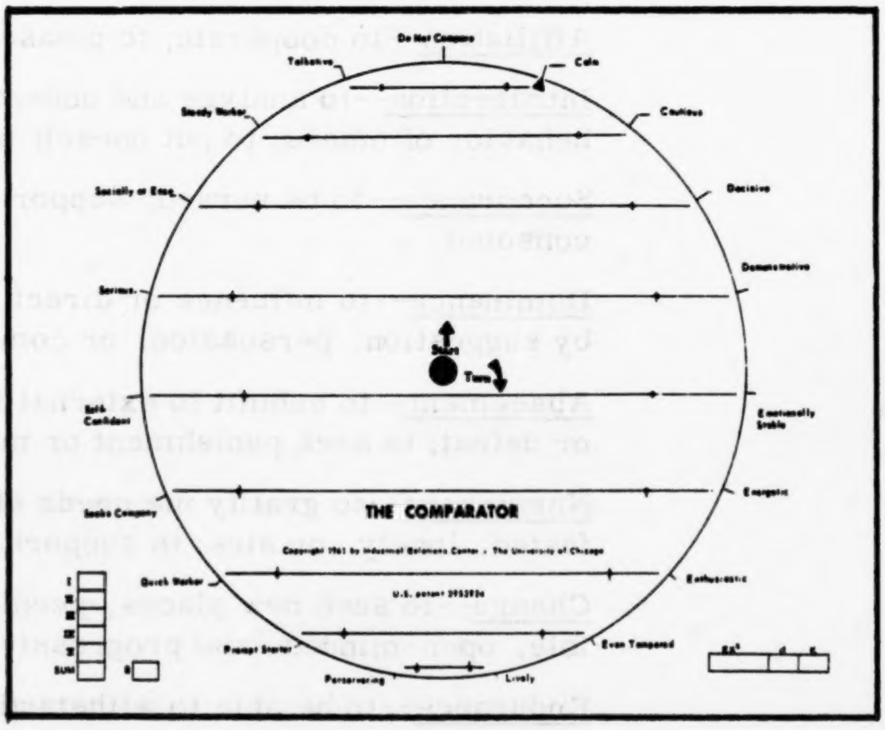
Circle with
Temperament Traits

SAMPLES:
(cont.)

Cardboard Disc


Start
Turn
THE COMPARATOR
Copyright 1961 by Industrial Relations Center - The University of Chicago
U.S. patent 2952924

Circle and Disc Assembled


Do Not Compare
Talkative
Calm
Steady Worker
Cautious
Socially at Ease
Decisive
Demonstrative
Serious
Start
Turn
Self Confident
Emotionally Stable
Energetic
Seeks Company
THE COMPARATOR
Copyright 1961 by Industrial Relations Center - The University of Chicago
U.S. patent 2952924
Quick Worker
Enthusiastic
Prompt Starter
Even Tempered
Persevering
Lively
SUM

TIME LIMITS: **Untimed. Approximately 15 to 20 minutes.**

EDWARDS PERSONAL PREFERENCE SCHEDULE [EPPS]

AUTHOR: Allen L. Edwards, University of Washington. Published by The Psychological Corporation, New York, N. Y.

MEASURES: This schedule serves as a measure of the relative strength in the individual of 15 "needs" adapted from Murray's basic need system. It yields a score for each of these needs, which are defined as follows:

Achievement--to accomplish something difficult; to rival and surpass others.

Deference--to admire and support a superior; to yield to the influence of another.

Order--to achieve organization, tidiness, and precision.

Exhibition--to make an impression; to entertain or shock others.

Autonomy--to resist coercion and restriction; to be independent and free to act according to impulse.

Affiliation--to cooperate; to please; to remain loyal.

Intraception--to analyze and understand the feelings and behavior of others; to put oneself in the place of others.

Succorance--to be nursed, supported, protected, guided, consoled.

Dominance--to influence or direct the behavior of others by suggestion, persuasion, or command.

Abasement--to submit to external force; to admit error or defeat; to seek punishment or misfortune.

Nurturance--to gratify the needs of the inexperienced, defeated, lonely, or sick; to support, protect, or heal.

Change--to seek new places, people, or ideas; to be flexible, open-minded, and progressive.

Endurance--to be able to withstand pain or hardship; to persist in an action despite opposition or setbacks.

Heterosexuality--to form and further erotic relationships with members of the opposite sex.

Aggression--to overcome opposition forcefully; to take revenge; to attack and injure others.

MEASURES:
(cont.)

The schedule itself consists of 210 pairs of statements in a forced choice format. The subject is asked to select the one from each pair which he considers better describes what he likes or feels.

SAMPLE ITEMS:

A I like to have my work organized and planned before beginning it.

B I like to travel and to see the country.

A I like to share things with my friends.

B I like to analyze my own motives and feelings.

TIME LIMITS: Untimed. Approximately 45 minutes.

THE PRESS TEST
[PRESS]

AUTHORS: Melany E. Baehr, Ph.D., and Raymond J. Corsini, Ph.D., Industrial Relations Center, The University of Chicago.

MEASURES:

This is an objective test of personality, developed to measure an individual's ability to work under stress. It yields three main scores, one from each of its three parts:

Part I.	Speed of reaction to verbal stimuli
Part II.	Speed of reaction to color stimuli
Part III.	Speed of reaction to color stimuli in a stress situation caused by the interference of distracting verbal stimuli.

Two further scores which can be derived are the figures for Part I minus Part II and for Part II minus Part III. Directions for each part, with sample items, are given below.

SAMPLE ITEMS:

Part I

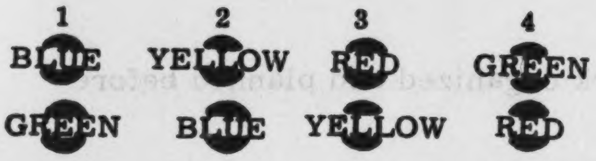
Put the first letter of each word in the corresponding circle.

1	2	3	4	1	2	3	4
Red	Green	Yellow	Blue	Ⓡ	Ⓖ	Ⓨ	Ⓑ
Yellow	Blue	Red	Green	Ⓨ	Ⓑ	Ⓡ	Ⓖ

SAMPLE ITEMS:
(cont.)

Part II

Put the first letter of the colors of the dots in the corresponding circles.



1	2	3	4
Ⓑ	Ⓨ	Ⓡ	Ⓖ
Ⓖ	Ⓑ	Ⓨ	Ⓡ

Part III

Write the first letter of the color of the ink in which the word is printed in the corresponding circles.

1	2	3	4	1	2	3	4
Red	Blue	Green	Yellow	Ⓡ	Ⓑ	Ⓖ	Ⓨ
Blue	Yellow	Red	Green	Ⓑ	Ⓨ	Ⓡ	Ⓖ

TIME LIMITS:

Part I 90 seconds.
Part II 90 seconds.
Part III 90 seconds.

BEHAVIOR INVENTORY
[BI]

AUTHORS: The McMurry Company, 645 N. Michigan, Chicago, Ill. Published by Industrial Relations Center, The University of Chicago.

MEASURES: In the Chicago Police Department Study, this inventory was referred to as the FTQ Scale or "Fail to Qualify" Scale. It is a very short form of the Minnesota Multiphasic Personality Inventory, and its 30 true-false items have been validated against investigations of personal background. It yields a single score which is predictive of whether or not an individual will successfully pass a personal-background investigation.

SAMPLE ITEMS:

In school I was sometimes sent to the principal for cutting up.	T F
I have never been in trouble with the law.	T F

TIME LIMITS:	Untimed. Approximately 10 minutes.

HAND TEST
[HAND]

AUTHOR:	Edwin E. Wagner, Ph.D. Published by The Mark James Co., Publishers, Akron, Ohio.
MEASURES:	The Hand Test is a projective test measuring an individual's characteristic modes of behavioral response to all facets of his environment. It yields eight basic scores:

Interpersonal Response--the tendency to relate to other individuals in an outgoing, active, and even aggressive manner, or else in a way that is receptive and inner-directed but not passive.

Environmental Response--the tendency to control, influence, or directly manipulate the external environment in order to alter objects or obtain goals.

Maladjustive Response--the unhealthy reaction of perceiving the environment as hostile and dangerous to one's physical well-being, with a corresponding concern about injury and disease.

Withdrawal Response--inability to meet the demands of the immediate environment, failure to adapt to it constructively and productively.

Inner-Directed Response--introversion, or the tendency to respond to others mainly in accord with one's own needs and wishes; undue concern with one's own mental state.

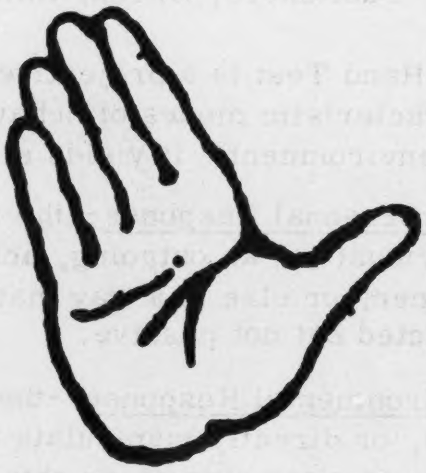
Outer-Directed Response--extroversion, or the tendency to direct behavior and interests strongly outward and to obtain satisfaction from things outside one's own self.

Acting Out Ratio--a measure of the degree to which an individual can absorb and control his own anti-social impulses; a predictor of the probability of maladaptive, overly aggressive behavior.

Pathological Score--a score which combines the Maladjustive and the Withdrawal Responses to the environment; an index of severe behavioral disturbance.

The test consists of nine pictures showing hands in various positions and a blank page where the subject is to supply a hand position of his own choice. The instructions are to look at each picture and then provide two descriptions of what the hand might be doing, such as "Waving to a friend ---a greeting" or "Policeman saying stop."

SAMPLE ITEM:



WHAT MIGHT THE HAND BE DOING?

____

____

TIME LIMITS: 8 minutes.

THE MAUDSLEY PERSONALITY INVENTORY [MPI]

AUTHOR: H. J. Eysenck, Ph.D., Institute of Psychiatry, University of London. Published by Educational and Industrial Testing Service, San Diego, California.

MEASURES: This inventory measures two relatively independent dimensions of personality or ranges of personality type: extroversion-introversion and neuroticism-stability. It consists of 48 questions to which the subject responds by marking "Yes," "?," or "No." For the Chicago Police Department study, the inventory yielded three scores:

Extroversion--extent to which the individual is outgoing, uninhibited, impulsive, and sociable.

Neuroticism--extent to which the individual is emotionally unstable, emotionally overresponsive, and liable to neurotic breakdown under stress.

Undecided--frequency with which the individual cannot decide on either the positive or the negative response to a question.

SAMPLE ITEMS:

Would you be very unhappy if you were prevented from making numerous social contacts? Yes ? No

Do you like work that requires considerable attention? Yes ? No

TIME LIMITS: Untimed. Approximately 10 minutes.

ARROW-DOT TEST
[A-D]

AUTHORS: Lawrence A. Dombrose, Ph.D., and Morton S. Slobin, Ph.D.

Published by Psychological Test Specialists, Missoula, Montana, and Industrial Relations Center, The University of Chicago.

MEASURES: This projective test is an adaptation of one part of a four-part battery called The IES Test. Its 30 items each set a relatively simple perceptual-motor task--to draw a line from the point of an arrow to a dot. Arrow and dot are separated by a number of barriers in the form of solid or dotted lines and black bars. The instructions explain how various combinations of these are to be handled. The test yields three scores:

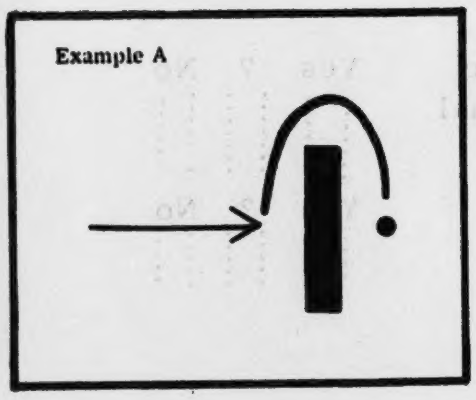
MEASURES: (cont.)

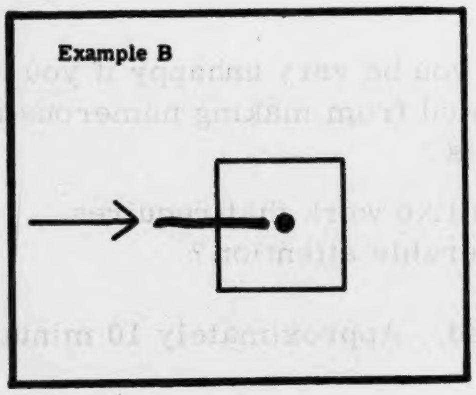
The I Score--a measure of the tendency to uncontrolled impulsive behavior through which needs are satisfied immediately and directly without any regard for realistic restrictions or moral issues.

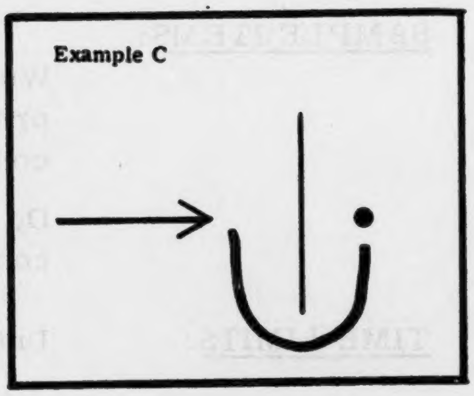
The E Score--essentially a measure of the ability to function realistically, of the degree of ego strength and control.

The S Score--a measure of superego strength, especially in circumstances where the demands of the superego are not experienced consciously but are viewed as external restrictions.

SAMPLE ITEMS:


Example A


Example B


Example C

TIME LIMITS: 6 minutes.

REFERENCES

Buros, Oscar Krisen (Ed.). The sixth mental measurements yearbook. Highland Park, N.J.: The Gryphon Press, 1965.

CHAPTER IV--APPENDIX

The Psychological and Psychiatric Evaluation Procedure Used in the Chicago Police Department

Initially all candidates are given a test of general intelligence (prepared by the Civil Service Commission) similar to the Army General Classification Test, and three other psychological instruments: a sentence completion test, a modified form of the Minnesota Multiphasic Personality Inventory, and a biographical form.

During this initial phase, applicants are eliminated only if they fail the I.Q. test (with a cut-off of an I.Q. of about 100). The other psychololological measures are used later to assist detectives during the background investigation which is conducted prior to acceptance into the Training Academy. This background investigation procedure, which is the second phase of processing consists of the following steps:

1. Fingerprint and name record
2. Employment history
3. Educational and military service records
4. Credit and neighborhood standing
5. Medical, psychological, psychiatric, and polygraph examinations
6. Personal family visits

During this phase, in rare instances, a man's background will be clear, but his psychological tests may indicate that he lacks the desired emotional stability. When this occurs, he will be evaluated by the Psychiatric Board (three psychiatrists) to determine his suitability for service as a police officer. This procedure is applied infrequently, since in nearly every instance the school, work, military, or criminal record confirm the questionable psychological pattern, and no difficulty is experienced in eliminating the unsatisfactory candidate.

Because of the large volume of applicants--2000-2500 candidates screened each year--it is impossible to use a polygraph or oral examination for each candidate. The Polygraph and Oral Boards are used to a limited extent when required on certain unusual cases. (The Oral Board, made up of psychologists, psychiatrists, and command personnel represents the ideal final screening device before employment.)

In phase three, when the successful candidates are admitted to the Academy, a battery of more sophisticated psychological tests is administered. Currently in use are: a Rorschach test, the Hand Test, the Picture Title test, and the Picture Memory test. In addition, a number of tests utilized in the present Department of Justice study are now routinely administered to Academy trainees, including the Personal History Index, Closure Speed, Closure Flexibility, Test of Social Insight, Press Test, Arrow-Dot Test, and Temperament Comparator. Recruits in the Academy are also given an Otis I.Q. test, the California Test of Mental Maturity, the Watson-Glaser test, and the Emo Questionnaire.

A number of these tests require interpretation and, in some cases, administration by trained psychologists. Personal interviews are conducted by the psychologists when testing indicates the existence of a problem area.

All this test material is evaluated by department psychologists, and their conclusions on each man are sent to the Director of Training and the Director of Personnel for analysis and possible action. When the probationer's psychological test results are unsatisfactory, or his actions deemed questionable during his three months in the Training Academy or remaining three months probation working as a patrolman, he is brought before the Psychiatric Board. This is seldom necessary because of the earlier elaborate screening described above.

CHAPTER IV--APPENDIX

Test Administration Procedures

Following the introduction of the staff members who would be administering the test battery, the participants were given a general orientation statement about the study. The statement was made in an informal manner rather than read, but the following paragraphs contain the principal points covered:

> "The purpose of this testing is to study the characteristics of successful police officers as an aid in the selection and placement of candidates for the position of patrolman. We are not searching for an ideal type of person. Many types of person can and do perform successfully as patrolmen.
>
> "The study is sponsored by the Office of Law Enforcement Assistance, U.S. Department of Justice. [At this point, Wave II subjects were told that this was the second half of the study, and that 250 Chicago patrolmen had already been tested.] Our work is not directly sponsored or paid for by the Chicago Police Department. This is very important for you, since your individual test results will not be made available to the Department. No test results from this study will be placed in your personnel folder, and our final report will be submitted to the Department of Justice. Your test performance here will have no bearing on your career as a police officer, in either a positive or a negative way. You are here on your own time, being paid by The University of Chicago.
>
> "We are trying to establish that these tests may have value for screening applicants for the patrolman's job. However, we would like you to take the tests as if they really counted. To insure meaningful results, we would like you to try and do your best.
>
> "All of you have probably had considerable exposure to testing, in the service as well as in your career as a police officer. From this point of view, you are fairly 'test sophisticated.' However, many of the tests that you will work through today may be different from anything you have seen before.

"Some of the tests have strict time limits, and it is very important that the group start and stop together. If someone starts too soon, or does not stop when the signal is given, those particular results will have to be discarded. In this situation, the person has taken the test under conditions different from those for the other members in the group, and his results would be invalid. The need for standard time limits is obvious, and much of your experience with testing has probably been under these conditions. On the other hand, some of the tests for today are 'untimed.' Their time limits are not strictly set, and all members of the group are allowed to complete the test. However, keep in mind that we have a fixed number of tests to complete, and the time when the group will finish depends on how long is taken for the untimed tests.

"A general rule to follow when you receive the test materials is 'Don't open the test booklet until told to do so!' Otherwise, your results on a particular test may be invalidated.

"We feel that the results of this study will have implications for police departments across the country. The findings we establish here in Chicago will be disseminated by the Department of Justice to the officials of all the larger departments and many of the smaller organizations. Thus your cooperation here today will have far-reaching repercussions for the professional advancement of police work."

Participants were encouraged to ask questions during this orientation statement and at its conclusion. Administrative details concerning rest breaks, the serving of coffee and rolls, the location of rest room facilities, and the arrangements for payment of the $15.00 volunteer's stipend were then discussed.

The administration procedure for each test followed the instructions contained in the particular test manual. The actual test administration began with Perceptual Speed, a timed test. The sequence of administration called for the alternation of timed and untimed tests to provide variety and a break from the pressure of the timed tests. There proved to be wide individual differences in the time required to complete particular untimed tests. In such instances, patrolmen who had completed their tests were allowed to leave the room for a brief rest until all members of the group were finished.

The last test in the four-hour session was one of the longer untimed personality tests. As patrolmen finished this test, they brought their materials to the front of the room, signed the payment roster, and received the $15.00 volunteer stipend in cash.

The actual length of the group sessions was very close to the planned four-hours' duration. A few men completed the test battery in less than four hours and most completed it in just a few minutes over this time, while a few required an additional 30 to 40 minutes to finish the last untimed test.

CHAPTER V

Results: Performance Appraisal

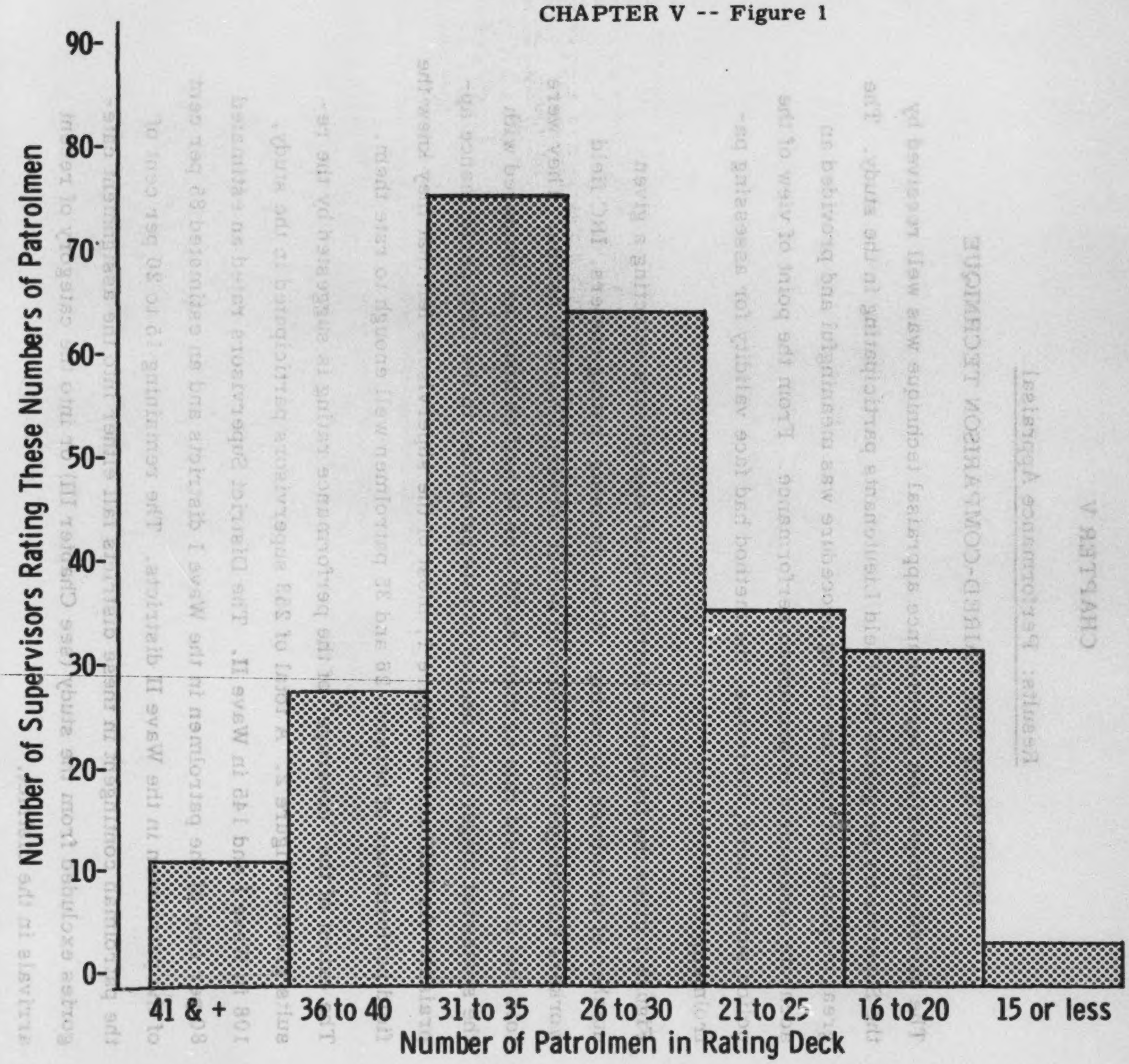
RESULTS OF THE PAIRED-COMPARISON TECHNIQUE

The paired-comparison performance appraisal technique was well received by the Supervisory Sergeants and Field Lieutenants participating in the study. The great majority felt that the rating procedure was meaningful and provided an accurate assessment of patrolman performance. From the point of view of the police supervisors, therefore, the method had face validity for assessing patrolmen.

Figure 1 gives the frequency distribution of supervisors selecting a given number of patrolmen for rating. When obtaining rating rosters, IRC field workers encouraged supervisors to select between 20 and 40 names. They were told that they could rate as many as 30 men in half an hour, as compared with the somewhat longer time needed to rate patrolmen in the CPD performance appraisal. As indicated in Figure 1, most of the supervisors felt that they knew the field performance of between 26 and 35 patrolmen well enough to rate them.

The extent of the "coverage" of the performance rating is suggested by the results given in Figure 2. A total of 253 supervisors participated in the study, 108 in Wave I and 145 in Wave II. The District Supervisors rated an estimated 80 per cent of the patrolmen in the Wave I districts and an estimated 85 per cent of the patrolmen in the Wave II districts. The remaining 15 to 20 per cent of the patrolman contingent in these districts fall either into the assignment categories excluded from the study (see Chapter III) or into the category of recent arrivals in the district.

CHAPTER V -- Figure 1



CHAPTER V -- Figure 2

SELECTION OF PATROLMAN SAMPLES FOR WAVE I AND WAVE II TEST ADMINISTRATION

WAVE I
DISTRICTS

		2	5	10	13	15	19	Total
I	No. in Watches	273	170	227	238	163	183	1254
II	No. Rated	173	143	166	191	131	143	947
III	No. Tested	44	29	40	42	19	68	242
IV	No. of Raters	21	17	19	15	15	21	108

WAVE II
DISTRICTS

		3	7	8	9	14	21	Total
I	No. in Watches	392	377	169	232	194	284	1648
II	No. Rated	316	294	158	192	174	246	1380
III	No. Tested	33	32	20	53	55	55	248*
IV	No. of Raters	27	24	21	23	26	24	145

* In addition to these district patrolmen, a total of 60 Task Force patrolmen were tested during Wave II. Their results are not given in this report, but will be detailed in a subsequent publication.

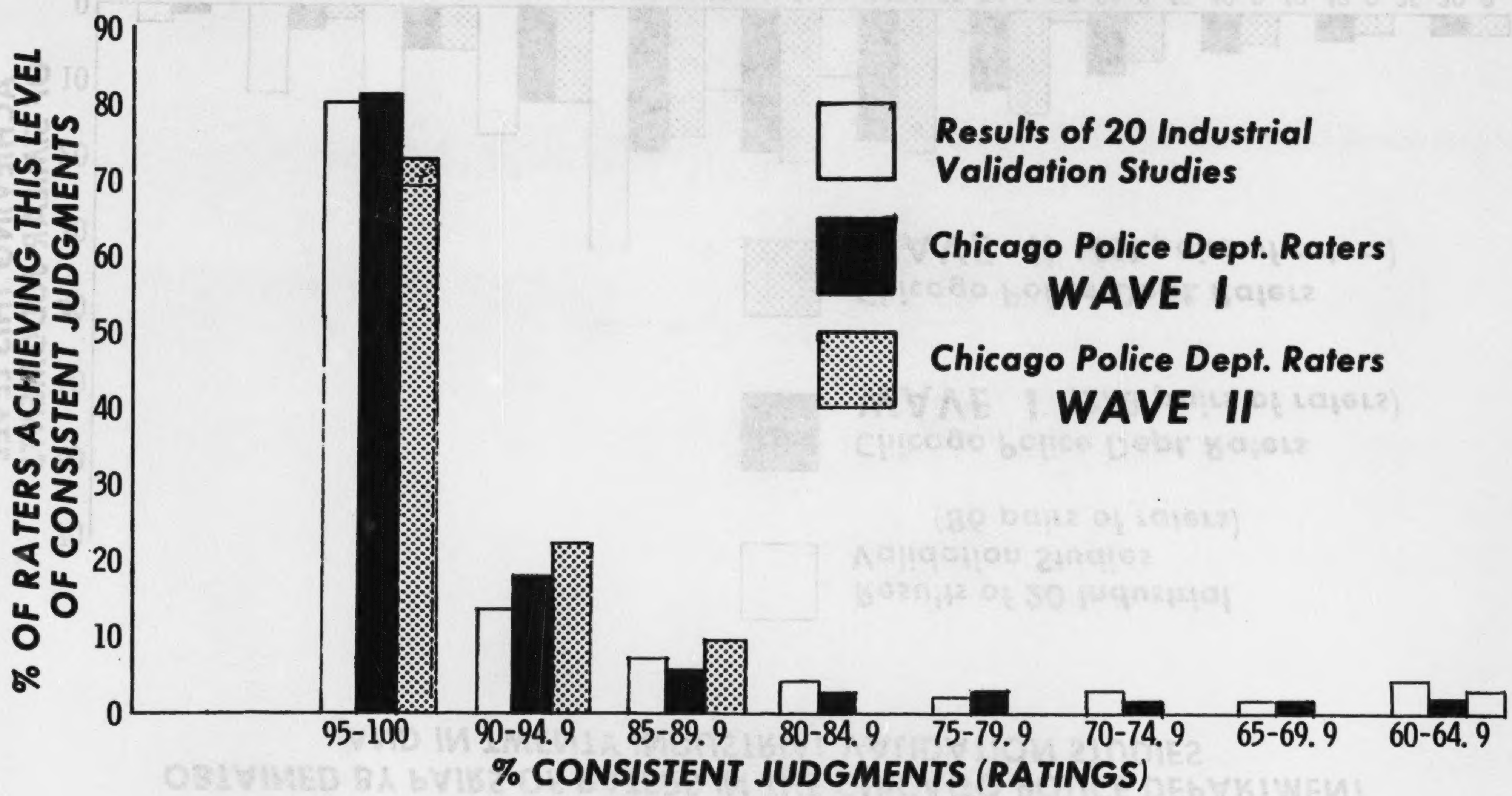
The distribution of the per cent consistent judgments is given in Figure 3 for the raters of Wave I and Wave II, as well as for raters in 20 industrial validation studies conducted by the IRC. The consistency of the judgments made in the study was quite high, with over 90 per cent of the police supervisors in both Waves achieving a consistency score of 90 per cent or more. These results compare quite favorably with the consistency results obtained in industrial studies. After a review of the results of the appraisal, raters with a consistency score lower than 80 per cent were ordinarily excluded from the study.

Figure 4 contains the results of the analysis of the agreement of raters within a watch. Once again, the results of Wave I and Wave II are contrasted with the results of the industrial validation studies. The results of the two testing Waves are quite comparable, but the police supervisor group exhibits a much wider range and a somewhat lower average level of agreement than does the industrial group. This difference may be the result, in part, of the generally larger size of the groups rated by CPD supervisors. Of greater consequence, however, are the closer day-to-day contacts which industrial supervisors ordinarily have with the employees they rate and the greater diversity and complexity of the patrolman's job as compared with many industrial occupations. The average level of agreement exhibited by industrial raters is in the 70 to 75 per cent range. The average level shown by the District Supervisors was in the 65 to 70 per cent range, which, under the circumstances, is regarded as a very favorable result. Police supervisors exhibiting an agreement index of less than 60 per cent were excluded from the study.

The performance index derived for each patrolman included in the study was therefore based on supervisory ratings which were internally consistent for each individual rater and also exhibited an acceptable degree of agreement among raters. Under these conditions, it was considered that the index would provide a reasonably accurate measure of patrolman performance and could serve as one of the major performance criteria for the study.

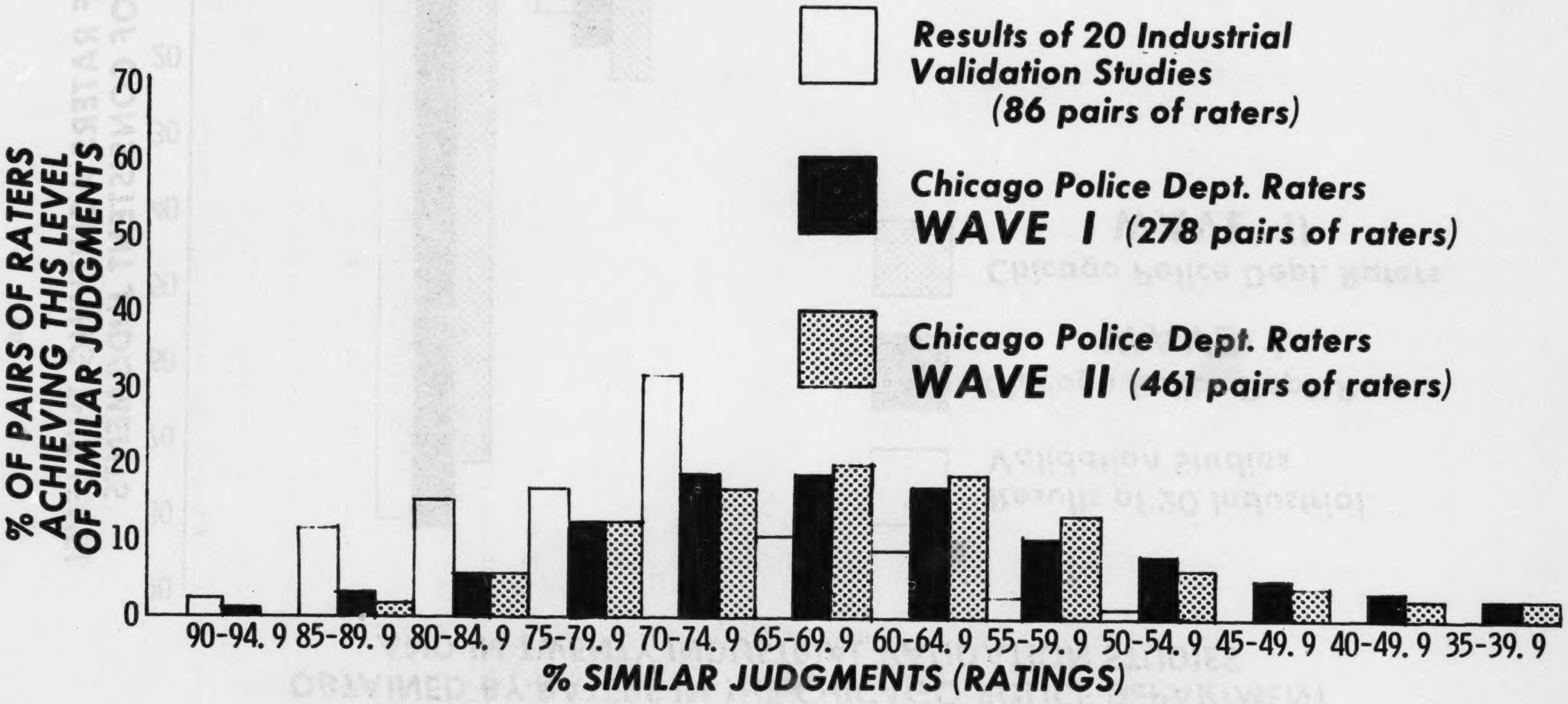
CHAPTER V -- Figure 3

COMPARISON OF PER CENT CONSISTENT JUDGMENTS
OBTAINED BY RATERS IN THE CHICAGO POLICE DEPARTMENT
AND IN TWENTY INDUSTRIAL VALIDATION STUDIES



CHAPTER V -- Figure 4

COMPARISON OF PER CENT SIMILAR JUDGMENTS OBTAINED BY PAIRS OF RATERS IN THE CHICAGO POLICE DEPARTMENT AND IN TWENTY INDUSTRIAL VALIDATION STUDIES



One of the wider implications of the successful implementation of the paired-comparison technique in the Chicago Police Department is the interchangeability of behavioral science procedures and techniques among different institutional settings. During the past two decades, the Industrial Relations Center has implemented test validation and performance appraisal procedures largely in industrial organizations. However, in recent years, these procedures have increasingly been applied in other institutions, such as schools, hospitals, and departments of the government. In general, the paired-comparison technique has been well received by raters and is economical of their time, since a rater can ordinarily assess 30 subjects in 30 to 40 minutes. In addition, it yields a variety of objective information concerning both the people rated and the raters. For these reasons, it should be seriously considered by law-enforcement agencies not only for specialized research purposes but also as part of the performance-review procedure aimed at the development of the individual.

RESULTS OF THE CALL FOR VOLUNTEERS

The results of the call for volunteers to participate in the study were deemed satisfactory. Eighty to 85 per cent of the patrolmen in the districts returned their participation forms, and of these approximately 60 per cent indicated willingness to participate. (It is estimated that approximately 10 per cent to 15 per cent of the complement of a district will be off duty during a given time interval because of furloughs, utilization of accumulated "time due," leaves of absence, illness, injury on duty, or suspension. Thus an 80 to 85 per cent return was the maximum anticipated.) Since the design of the study called for a carefully designed sample composed of some of the better and some of the poorer performers from each district, ample representation of these ranges of performance was available within the volunteer group.

Better than 90 per cent of the patrolmen invited to participate in the study actually attended the group testing sessions. As indicated previously, the tests were administered in two "Waves," Wave I during February of 1967 and Wave II during

July of 1967. The testing was scheduled in this way to allow sufficient time for the conduct of the paired-comparison performance appraisal in the districts and for analysis of Wave I test results so that the test battery could be refined, if necessary, for use during Wave II.

A total of 490 district patrolmen were tested in the study, as indicated in the district-by-district breakdown contained in Figure 2. Participants were selected and notified as described in Chapter III. The sessions were conducted by the IRC staff at two locations in the city, using the classroom facilities of DePaul University on the North side, and The University of Chicago Law School on the South side. The actual test administration procedures have been described in the Appendix to Chapter IV.

RESULTS OF CHICAGO-POLICE-DEPARTMENT APPRAISAL MEASURES

In implementing selection validation procedures in organizations, the Industrial Relations Center's policy is to use both subjective and objective indices of performance available in the organization in addition to those obtained from the paired-comparison technique. This has the double advantage of affording a comparison between the results obtained from the paired-comparison technique and from the appraisal indices used by the organization, and of investigating test validity for objective, operational indices regarded as important by the organization.

The second major criterion measure used in the study was the supervisory performance rating, conducted semi-annually in the department. This subjective measure and the six objective performance indices were described in Chapter III. The intercorrelations of these measures are presented in Table 1. Intercorrelations based on Wave I data are presented above the diagonal. Wave II intercorrelations are presented below the diagonal. These intercorrelations are generally consistent across the two Waves, both in magnitude and in direction. The highest correlations and hence greatest overlap occur between the paired-comparison and CPD ratings. As discussed in Chapter III, both indices are

CHAPTER V -- Table 1

Intercorrelations of the Eight Major Patrolman Performance Criteria for Wave I (N=175) and Wave II (N=192)*

Variable	21	372	308	373	375	374	359	314
PC Rating (21)	X	.64	.12	.33	.08	-.19	.36	-.14
CPD Rating (372)	.66	X	.17	.23	-.07	-.28	.06	-.07
CPD Tenure (308)	.16	.22	X	-.23	.10	.08	-.01	.03
CPD Awards (373)	.27	.24	.07	X	.11	-.04	.11	-.05
IID Complaints (375)	.02	.08	.13	.27	X	.29	.42	.02
Disciplinary Actions (374)	-.10	-.04	.34	.18	.48	X	.09	.09
No. of Arrests (359)	.19	.21	.05	.48	.23	.23	X	-.08
Times Absent (314)	-.06	-.04	.20	.01	.12	.34	.01	X

*** Wave I intercorrelations are presented above and Wave II intercorrelations below the diagonal line. For 175 cases, r of .17 or greater is statistically significant at the .01 level, and an r of .12 or greater is statistically significant at the .05 level.**

global measures of patrolman performance, and a fairly high (but not perfect) correlation would be expected between them. These results evidence both the basic soundness of the CPD rating, and the utility and validity of the paired-comparison rating. Since the overlap is considerable but not outstanding, the results are thought to support the contention made in Chapter III that the paired-comparison and CPD ratings assess patrolman performance in this study from slightly different points of view--field performance vs. field as well as administrative performance.

As might be expected from their high intercorrelation, the paired-comparison and CPD ratings show similar relationships with the six objective measures. Both show a significantly positive but not striking relationship with tenure, and a more substantial relationship with CPD awards. There is virtually no relationship between the results of the two supervisory ratings and either the IID-complaints or times-absent criteria. A negative relationship exists between the two ratings and disciplinary actions, but these correlations are significant only in the Wave I results. The total-number-of-arrests index is positively related to both the paired-comparison and CPD ratings.

The relationship between CPD tenure and the remaining indices is more mixed, with a negative relationship with awards in Wave I, and positive relationships with complaints, disciplinary actions, and times absent in Wave II. Departmental awards do not correlate significantly with any of the remaining Wave I indices, but show positive relationships with complaints and disciplinary actions, and a strong positive relationship with total number of arrests. Complaints are significantly related to disciplinary actions and arrest total in both Wave I and Wave II. In Wave II, disciplinary actions are significantly related to both arrests and times absent. The general direction and strength of associations met our expectations, and were thought to verify the anticipated relationships between some of the indices and the expected independence of others.

CHAPTER VI

Validation of the Occupational Test Batteries

The basic question to be answered in the validation of the occupational test batteries is to what extent the test (predictor) variables can account for differences in the performance (criterion) variables. To achieve a significant relationship between these predictor and criterion measures, the tests used in the battery should yield reliable measures of some of the personal attributes which are essential for successful performance and the criterion measures should reliably assess this performance. The basis for the selection of the tests is described in Chapter II, which analyzes the patrolman's working environment; the actual tests selected and the criteria for selection are described in Chapter IV. Equal care was exercised in the selection of the criterion measures. An appraisal of patrolman performance, using the paired-comparison technique, was specifically implemented for this study, and its results are described in Chapter V. In addition, seven other criterion measures were drawn from available records kept by the Chicago Police Department. The most important of these were the regularly implemented departmental appraisal and tenure. All of the criterion measures are described in Chapter III.

Before the basic question of the predictability of these performance measures could be answered, two subsidiary questions relating to the experimental design had to be dealt with. Two separate samples of patrolmen, called the Wave I and the Wave II samples, were used in this study; the basis for their selection is described in Chapter III. A common core of tests was administered to both samples. However, in order to investigate as many predictor variables as

possible within the necessary limitations of time, some tests used for the Wave I sample were dropped and others substituted for the Wave II sample. One question, therefore, was whether the contribution of Wave-specific tests could improve predictions based only on Wave-common tests. If it could, then Wave-specific validations were in order. The second question in relation to experimental design concerned the validation technique which would be used. In making this decision, the fact that both samples in this study were composed of a racially-mixed group of subjects had to be considered.

Should we succeed in selecting an appropriate validation technique and in demonstrating a significant degree of association between the test and the performance measures, we shall have reached the statistical objective of the validation. Thereafter, it is necessary to identify those areas of measurement and the specific test variables which contribute most heavily to the prediction. This will establish the groundwork for a predictive validation study, which should ordinarily follow the concurrent validation model employed here. Furthermore, for purposes of test interpretation, currently employed patrolmen should be described in terms of the predictive test variables identified in the validation. The first approach to achieving these objectives and to answering the subsidiary questions posed in this study was through a preliminary multiple-regression analysis.

THE PRELIMINARY MULTIPLE-REGRESSION ANALYSIS

The objective of the multiple-regression analysis is to select those test variables --together with the proper weights which should be applied to them, called beta-weights--which will best predict the criteria. Thereafter, the prediction for any individual is made by summing his properly weighted scores for each selected test variable. In this analysis, the Wave I and Wave II samples were treated separately. In addition, each sample was subdivided by race into white and Negro groups.

Multiple regressions on each of the eight criteria were calculated twice for each subsample. The first analysis considered all possible variables for inclusion in the model, while the second was restricted to those variables common to both Waves. In each analysis, the test variable making the largest contribution to the criterion variance was selected first for inclusion in the model. Its effect was then partialled from the remaining variables, and the next largest contributor was added. This selection process continued until the multiple correlation corrected for statistical bias (R') failed to increase with a newly included variable. The results of this preliminary analysis are summarized in Table 1.

The coefficients of multiple correlation (R) indicate the strength of the relationship between the observed criterion measures and a prediction of those measures based on the proper weighting of the test variables. The number of variables (m) used to generate the prediction and the estimate of R corrected for statistical bias (R') are also given. A further explanation of these measures may be found in Chapter 16 of Guilford (1956).

Table 1 shows that for both the Wave I and the Wave II samples the magnitudes of R' indicate significant relationships across the board for all criteria contributing to the study. In general, however, the highest corrected correlations (in the neighborhood of .6) were obtained for the first three criterion measures (the paired-comparison rating, the CPD performance rating, and tenure) which are regarded as the major criterion measures of the study. These results encourage us to believe that the tests selected for this study are significantly related to many of the important aspects of police patrolman performance. Given the magnitudes of R', it should also be possible to select a smaller set of test variables which will effectively predict the criterion measures. Of further interest is the fact that the magnitudes of R' based on tests common to both Waves tend to be smaller than those based on Wave-specific tests. This discrepancy is more evident for the Wave II sample than for the Wave I sample, suggesting that the variables unique to Wave II may be important for police selection. In any event,

CHAPTER VI -- Table 1

Multiple Correlations of Test Batteries with Criterion Measures

TEST BATTERIES	Groups	N	CRITERION MEASURES																							
			PC Rating			1966 CPD Rating			CPD Tenure			CPD Awards			IID Complaints			Disciplinary Actions			Number of Arrests 1966			Times Absent		
			R	R'	m	R	R'	m	R	R'	m	R	R'	m	R	R'	m	R	R'	m	R	R'	m	R	R'	m
WAVE I SAMPLE																										
Wave I Tests	Total	175	.60	.50	25	.67	.61	23	.72	.67	21	.57	.48	20	.53	.46	14	.47	.38	16	.53	.46	15	.55	.43	25
	White	128	.62	.53	19	.66	.57	22	.74	.67	22	.59	.47	20	.64	.53	24	.51	.38	18	.63	.51	24	.67	.55	26
	Negro	47	.92	.85	22	.90	.83	17	.92	.88	13	.85	.74	18	.91	.82	21	.89	.75	24	.84	.71	19	.96	.83	33
Tests in Common to Both Samples	Total	175	.58	.49	23	.65	.59	20	.67	.63	16	.47	.39	14	.48	.41	14	.44	.36	13	.45	.38	12	.50	.40	18
	White	128	.61	.51	20	.63	.56	16	.68	.62	16	.47	.37	12	.58	.46	19	.43	.33	11	.54	.47	12	.59	.49	19
	Negro	47	.90	.80	21	.94	.87	24	.99	.96	33	.86	.75	18	.92	.80	27	.93	.85	23	.94	.86	24	.85	.73	18
WAVE II SAMPLE																										
Wave II Tests	Total	192	.67	.60	27	.63	.55	27	.81	.78	19	.57	.50	20	.61	.52	28	.66	.57	33	.54	.47	19	.60	.50	26
	White	126	.77	.65	37	.69	.56	29	.84	.81	16	.60	.53	15	.65	.53	24	.77	.66	33	.55	.44	17	.74	.63	32
	Negro	66	.98	.93	41	.88	.79	26	.97	.93	37	.88	.79	28	.95	.89	36	.88	.77	29	.88	.80	24	.95	.86	43
Tests in Common to Both Samples	Total	192	.54	.48	16	.50	.43	17	.80	.78	15	.49	.41	17	.51	.43	16	.58	.51	19	.44	.37	12	.38	.29	13
	White	126	.67	.57	24	.59	.48	19	.84	.81	14	.45	.36	11	.57	.50	13	.65	.57	19	.44	.33	12	.50	.41	13
	Negro	66	.81	.73	18	.93	.85	35	.92	.88	20	.87	.79	25	.85	.76	23	.86	.75	26	.74	.62	17	.75	.60	21

this result prompted the decision to continue the multiple-regression analyses separately for each Wave.

It should also be noted that for both the Wave I and the Wave II samples, and for both the Wave-specific batteries and the tests common to both Waves, the multiple correlations obtained for separate racial groups are higher than that obtained for the racially-mixed total group. For example, for the paired-comparison performance rating, the corrected multiple correlations for the total, white, and Negro groups in the Wave I sample are .60, .62, and .92 respectively. In the Wave II sample the corresponding coefficients are .67, .77, and .98. Table 1 shows that the most spectacular increases and the greatest absolute values of R' for all criteria studied were obtained for the Negro group. These results led to the decision to continue the two Wave analyses separately for each racial group. Further results of these separate racial validations, together with a discussion of the recommendations for test usage made by the Equal Employment Opportunity Commission and also of the implications for further recruitment and selection, are presented in detail in the following Chapter.

THE PRINCIPAL MULTIPLE-REGRESSION ANALYSIS

The decision to partition the data by Waves and by racial groups forced a revision in the proposed two-way cross-validation design. We had originally intended to determine test variable beta-weights on the criteria for each Wave separately and then apply these weights to the other Wave. The sample used to derive beta-weights is considered the primary validation sample. The other sample to which those weights are applied is referred to as the cross-validation sample. The correlations between the observed criteria and the predictions based on the betas from the other Wave would have represented cross-validation indices and would have shown how well the predictions fared across samples.

The major revision to this design was a result of the redefinition of the samples. The members of each Wave were randomly assigned to one of two parts, and each part was then divided into total, white, and Negro subsamples. Hence, there were six subsamples per Wave, and, as a result, each Wave could provide its own primary and cross-validation samples. For example, within each Wave, a set of beta-weights was first generated using data from the Part 1, White sample and then used to predict the criterion in the Part 2, White sample. In this case, the Part 1 group would represent the primary validation sample while the Part 2 group would be the cross-validation sample. The correlation between the prediction and the observed criterion value in the Part 1 data is equivalent to R and is sometimes called the primary validation index. The same correlation, when generated for the Part 2 data, can be called the cross-validation index. This two-way feature is retained, since we treat the Part 2 sample as a primary validation sample, derive the beta-weights within it, and apply these to the Part 1 sample. In addition, this procedure is repeated in the given Wave for the Negro subsamples, and for the total sample. Finally, the entire operation is carried out in the other Wave. To accomodate this expanded design, the number of criteria treated was reduced. It was felt that the paired-comparison rating, the 1966 CPD rating, and CPD tenure represented the broad performance measures pertinent to this study and that the other five more-narrowly-defined measures could be dropped from consideration without loss of generality.

It was also decided to use more stringent controls in selecting variables for the principal regression analysis. The previously described preliminary regression procedure selected an unwieldy number of predictors, some of which appeared to be adding very little to the model while others appeared to be acting as suppressors. Both situations are unstable for cross validation.

The major considerations in selecting predictors are significance and directionality. "Significance" refers to whether or not the magnitude of the simple correlation (r) between any test variable and the criterion statistically exceeds zero in a primary sample. These values of r depend on the sample size and can be found in tables such as those in Guilford (1956). "Directionality" refers to whether or not a test variable which had a significant r in the primary sample carried the same sign in the cross sample, irrespective of magnitude. Hence, a variable was selected for inclusion in the regression analysis if its correlation with the criterion was significantly different from zero in the primary sample and did not change its sign in the cross-validation sample. All variables thus selected for any criterion in any primary sample were forced into the regression model, the beta-weights were calculated, and the regression equation applied to the cross-validation sample. The correlation between observation and prediction then represented the cross-validation index.

Some investigators will consider that this procedure for the selection of the variables introduces spurious effects into the cross-validation coefficients. However, there seemed to be some good reasons for pursuing this approach. The data collected for the study restrict the approach to validation of concurrent validation models. This means that any beta-weights derived from these data should be expected only to predict current performance in new samples. Of course, it may be of little practical value to use tests to predict current performance, since this can be more easily appraised by direct measurement. However, it would be of immense value to be able to predict a person's future performance on the basis of present test results. If a variable shows consistent and significant correlations with a criterion across four separate samples from the population of Chicago patrolmen, it seems reasonable to assume that this variable will show significant correlations with other samples of patrolmen. The identification of groups of variables showing such criterion correlations should provide information on the areas of measurement covered in the battery and the types of test which could be most fruitfully utilized in a predictive validation. Since the identification of such variables was one of the major objectives of this exploratory study, the approach taken was considered to be of greater ultimate benefit than the classically pure cross-validation procedure. The results of this analysis are summarized in Tables 2 through 7. These Tables present the information on the multiple correlation and the predictor variables and an approximation of the cross validations. In the first section, the designation of the primary validation sample heads each column.

Immediately beneath this information is the size of the sample (N), the number of predictor variables (m), the multiple correlation (R), and the R corrected for statistical bias (R'). The predictor section presents the simple correlation (r) of predictors with the specified criterion and their beta-weights (β) as calculated for that primary sample. Wave-specific predictors are marked by an asterisk. For the cross-validation section, the weights in the columns are applied to data from the various samples indicated by the row headings. For the cross-validation section of these Tables, the numbers presented are the correlations between predictions and observations of the criteria. Again, the predictions are made by applying the beta-weights derived from the data of the oolumn to the data named in the row heading. The sets of four correlations enclosed in the boxes represent the usual validation and cross-validation coefficients with the restrictions imposed on the latter as discussed earlier. The correlations based on beta-weights derived and applied to the same sample represent primary validation indices and should agree with the R at the head of the column. The other two numbers in each box show the results when weights derived in one sample are used to generate predictions in another sample. The chief focus of this Chapter is the within-Wave, within-race results. The latter regressions, in general, produce the highest cross-validity coefficients, as shown by the additional information given in the Tables for comparative purposes. This fact tends to support our earlier decision to continue separate validations for each subgroup.

IDENTIFICATION OF TEST VARIABLES MOST LIKELY TO PREDICT PATROLMAN PERFORMANCE

The first step in analyzing the results presented in the Tables is to examine the multiple correlations. The magnitude of R will indicate how well these variables combine to predict the criteria. Secondly, the cross-validation results provide information as to how far such predictions can be generalized across different samples. Once the predictions generated are thus evaluated, the relative importance of the predictor variables for the prediction of each criterion can be determined. There are several considerations in making this determination. A set of predictors with high primary and cross-validation indices may be ranked by the absolute value of their beta-weights. However, since there are several sets of predictors for each criterion, other rules must aid in this ranking. It seems reasonable to attribute predictive importance to a variable that occurs in several of the data sets available for that criterion. In addition, we may state rules that would decrease the value of a variable used for predictions. If a variable has a beta-weight opposite in sign to the correlation shown, that variable may be considered to be a suppressor and not a predictor. In other words, it is aiding prediction not by adding but by suppressing or subtracting variance from the regression model. It has been our experience that although suppressors aid in the primary validations, they are unstable between samples and so hinder effective cross-validation. Hence, suppressors as well as predictors with beta-weights of less than .10 will not be considered in this discussion unless they occur consistently across different data sets.

PREDICTION OF THE PAIRED-COMPARISON RATING

Tables 2 and 3 show that the area of background and experience, as measured by variables from the Personal History Index, contributes heavily to the prediction of this criterion. The variable Early Family Responsibility appears in a positive direction in three of the four possible within-race regressions for the Wave I sample and may thus be regarded as relevant to both racial groups. The only other variable in this area with a possibility of overlap between the racial groups is the negative prediction of Selling Experience, which occurs once each for the Wave I and Wave II samples. The remaining variables are race-specific and, in the case of the white group, appear only in one or the other of the two possible race-specific, within-Wave regressions. One would thus place less confidence in their predictive validity. However, for the sake of completeness they are listed here for the white group as positive weights for Parental Family Adjustment and Professional Successful Parents and a negative weight for School

CHAPTER VI—Table 2

Validation and Cross-Validation Results for PC Rating within Wave I
(total N = 181)

MULTIPLE CORRELATION		PART GROUP	1 TOTAL		2 TOTAL		1 WHITE		2 WHITE		1 NEGRO		2 NEGRO	
		N		80		101		58		73		22		28
		m		5		4		7		8		7		6
		R		.38		.41		.59		.53		.73		.66
		R′		.30		.36		.50		.44		.55		.52
	PREDICTORS	var #	r	β	r	β	r	β	r	β	r	β	r	β
PHI	1 School Achievement	1							-.21	-.15				
	5 Financial Responsibility	5									-.32	-.50	-.30	-.30
	6 Early Family Responsibility	6	.24	.19			.20	.17	.24	.22	.33	.00		
	13 Vocational Satisfaction	13											-.35	-.40
	14 Selling Experience	14	-.19	-.18			-.25	-.32						
	15 General Health	15									.34	.17		
AC	II Reasoning	23*					.19	.30						
TC	IV Excitable	34											-.33	-.03
	V Sociable	35	-.19	-.11							-.34	-.31		
	b Cautious	37					-.22	-.28						
	g Enthusiastic	42											-.28	-.22
HAND	Interpersonal Response	70*									.34	.16		
	Pathological Score	83*							-.20	-.15				
MPI	Undecided	86*					-.24	-.21						
SI	Withdrawal	88			-.19	-.08			-.19	-.07	-.32	-.16		
	Passivity	89	-.17	-.00					-.18	.03				
	Cooperation	90	.23	.20	.26	.23	.20	.11	.24	.20	.33	.31	.30	.29
	Competition	92					-.23	-.18						
A-D	I Score	96			-.24	-.25			-.24	-.27				
	Number of Responses	99			-.20	-.16			-.18	-.25			-.29	-.15
		*Unique to Wave I												
CROSS-VALIDATION	PART	GROUP												
	1	Total		.38		.23		.36		.21		.22		.19
	2			.27		.41		.24		.47		.26		.29
	1	White		.37		.17		.59		.21		.05		.12
	2			.33		.42		.21		.53		.19		.19
	1	Negro		-.41		.44		-.02		.25		.73		.36
	2			.14		.40		.29		.37		.46		.66

CHAPTER VI—Table 3

Validation and Cross-Validation Results for PC Rating within Wave II
(total N = 192)

MULTIPLE CORRELATION		PART GROUP	1 TOTAL		2 TOTAL		1 WHITE		2 WHITE		1 NEGRO		2 NEGRO	
		N	97		95		61		65		36		30	
		m	6		5		9		7		7		7	
		R	.49		.46		.59		.61		.66		.75	
		R′	.43		.41		.49		.54		.54		.65	
PREDICTORS		var #	r	β	r	β	r	β	r	β	r	β	r	β
PHI	1 School Achievement	1	-.27	-.21										
	4 Leadership and Group Participation	4									-.32	.10	-.36	-.23
	6 Early Family Responsibility	6					.20	.14						
	7 Parental Family Adjustment	7	.22	.15			.37	.11						
	9 School Activities	9									-.42	-.18		
	10 Professional-Successful Parents	10					.28	.26						
	14 Selling Experience	14									-.36	.05		
CLOSURE SPEED		18									-.38	-.19		
TC	III Self-Reliant	33					.20	-.26						
	c Decisive	38					.28	.24						
	i Lively	44					-.24	-.24						
	j Persevering	45											-.44	-.10
	n Self-Confident	49			.22	.27	.31	.21					.44	.21
SI	Competition	92			-.27	-.32			-.34	-.44				
	Total	95			-.23	.04	-.21	-.22	-.26	.17				
WII	4 Artistic and Interpretative	644*	-.22	-.16							-.43	-.25		
	Aspiration Level	654*							-.20	-.24				
CREE	Total	656*							.20	.28				
	2 Indifference vs. Involvement	658*					-.20	-.17					.43	.23
	5 Unsystematic, Selective vs. Systematic, Prescribed Activity	661*											-.37	-.19
	12 High vs. Low Artistic	668*	-.28	-.18							-.48	-.24	-.34	-.10
EPPS	Deference	671*	.25	.15							.39	.29	.36	.37
	Autonomy	674*	-.22	-.12	-.21	-.19			-.26	-.21				
	Nurturance	680*			.23	.17			.21	.19				
	Aggression	684*							-.20	-.36				

*Unique to Wave II

CROSS-VALIDATION PART	GROUP						
1	Total	.49	.25	.35	.05	.46	.32
2		.24	.46	.19	.40	.01	.12
1	White	.36	.33	.59	.12	.25	.22
2		.11	.48	.24	.61	-.14	-.17
1	Negro	.57	.15	-.05	-.11	.66	.51
2		.49	.46	.04	.08	.34	.75

Achievement. For the Negro group, we have the rather anomalous situation of negative weights for Financial Responsibility in two regressions and negative correlations for Leadership & Group Participation in two regressions, but the beta-weights have inconsistent signs. In addition, one regression for a Negro group shows a negative weight on School Activities and Vocational Satisfaction, and a positive weight on General Health.

Another set of variables which contribute heavily to the prediction of the paired-comparison rating are those derived from the Test of Social Insight. The variable Cooperativeness predicts positively in two sets of both white and Negro regressions and, consistent with this, the variable Withdrawal predicts negatively and occurs in one regression for each racial group. In addition to this, Competitiveness appears in two white regressions only and predicts negatively.

In the area of behavior attributes as measured by the Temperament Comparator, the temperament trait Self-Confident predicts in a positive direction in both a white and a Negro regression. The remaining variables are race-specific, with the Negro groups showing negative weights on such traits as Enthusiastic and Persevering and on the factor Excitable, and the white group showing negative weights on the traits Cautious and Lively but positive on Decisive.

On the Arrow-Dot Test, the total number of responses predicts negatively on both white and Negro regressions and the Id dimension also predicts negatively for the white group.

Of the Wave-specific tests, those for Wave II tend to contribute more to prediction. These are the Edwards Personal Preference Schedule, the Cree Questionnaire, and the Work Interest Index, but these results will not be discussed in detail although they are given in the Tables.

PREDICTION OF THE CHICAGO-POLICE-DEPARTMENT RATING

It will be seen from Tables 4 and 5 that the predictor variables which emerged for this criterion range across all the areas of measurement in the battery. However, the most fruitful area from which to select predictors appears, again,

CHAPTER VI—Table 4

Validation and Cross-Validation Results for CPD Rating within Wave I

MULTIPLE CORRELATION		PART	1		2		1		2		1		2	
		GROUP	TOTAL		TOTAL		WHITE		WHITE		NEGRO		NEGRO	
		N		80		101		58		73		22		28
		m		11		4		9		6		5		7
		R		.63		.53		.61		.54		.65		.71
		R′		.54		.50		.50		.48		.49		.58
	PREDICTORS	var #	r	β	r	β	r	β	r	β	r	β	r	β
PHI	2 Higher Educational Achievement	2					.17	.20						
	3 Drive	3			.21	.32								
	5 Financial Responsibility	5											-.41	-.32
	6 Early Family Responsibility	6	.38	.31			.37	.33	.20	.13				
	9 School Activities	9							.22	.22				
	12 Vocational Decisiveness	12											.29	.29
CLOSURE SPEED		18									.43	.14		
AC	II Reasoning	23*											.34	.28
PRESS	II-III	29	.21	.10			.20	.21						
TC	I Controlled	31											.38	.15
	V Sociable	35	-.22	-.23			-.21	-.06						
	d Demonstrative	39	-.19	-.21	-.33	-.31			-.34	-.22			-.27	.01
	k Prompt Starter	46	.19	-.16	.32	.26	.25	.11					.31	.20
	m Seeks Company	48							-.11	.04				
	n Self-Confident	49							.25	.22				
	p Socially At Ease	51									-.42	-.69		
	q Steady Worker	52			.28	.11			.37	.25				
HAND	Acting Out Ratio	80*											.40	.20
	Pathological Score	83*	-.19	-.19			-.20	-.25						
BEHAVIOR INVENTORY		87*					-.18	-.03						
SI	Withdrawal	88	-.23	-.18										
	Passivity	89	-.25	-.18			-.19	-.10			-.35	-.34		
	Cooperation	90	.32	.08			.26	.22						
A-D	E Score	97	.20	.11							.50	.21		
	S Score	98	-.28	-.08							-.51	-.39		

*Unique to Wave I

CROSS-VALIDATION PART	GROUP	1 TOTAL	2 TOTAL	1 WHITE	2 WHITE	1 NEGRO	2 NEGRO
1	Total	.63	.20	.52	.24	.45	.07
2		.19	.53	.18	.39	.02	.37
1	White	.54	.10	.61	.21	.24	.01
2		.25	.53	.28	.54	-.01	.23
1	Negro	.67	.25	.38	.19	.65	.35
2		-.03	.52	-.10	-.03	.11	.71

CHAPTER VI—Table 5

Validation and Cross-Validation Results for CPD Rating within Wave II

MULTIPLE CORRELATION	PART GROUP	1 TOTAL		2 TOTAL		1 WHITE		2 WHITE		1 NEGRO		2 NEGRO	
	N	97		95		61		65		36		30	
	m	6		5		10		5		11		6	
	R	.54		.44		.62		.39		.80		.74	
	R′	.49		.39		.51		.29		.70		.65	
PREDICTORS	var #	r	β	r	β	r	β	r	β	r	β	r	β
PHI 1 School Achievement	1	-.23	-.21										
5 Financial Responsibility	5											.52	.32
8 Stability	8	.28	.25	.23	.20	.20	.17			.40	.21		
10 Professional-Successful Parents	10					.23	.25						
15 General Health	15					.21	.20						
CLOSURE SPEED	18					-.26	.02						
PERCEPTUAL SPEED	20					-.24	-.20						
PRESS I-II	28	-.22	-.19			-.31	-.13						
TC j Persevering	45					-.26	-.25						
n Self-Confident	49			.21	.22								
SI Passivity	89							-.20	-.07	.35	.26		
Cooperation	90			.20	.19	.20	.11	.26	.12				
Competition	92							-.29	-.38				
Aggression	94									-.36	-.20		
TOTAL	95							-.20	.17				
WII 3 Authority & Prestige	643*									-.33	-.09		
4 Artistic and Interpretative	644*	-.31	-.23							-.51	-.03		
5 Artistic and Stylized	645*									-.44	-.22		
6 Artistic and Creative	646*	-.30	-.15			-.20	-.12			-.38	-.00		
11 Mechanical & Productive	651*									-.40	-.21		
CREE TOTAL	656*							.22	.16				
1 Dominance vs. Submission	657*									-.38	-.14		
12 High vs. Low Artistic	668*	-.27	-.04							-.37	.04	-.35	-.22
EPPS Achievement	670*			-.21	-.19							-.41	-.20
Deference	671*									.33	.41	.43	.17
Nurturance	680*			.20	.13							.43	.29
Endurance	682*					.22	.24						
Heterosexuality	683*											-.35	-.06

*Unique to Wave II

CROSS-VALIDATION PART	GROUP	1 TOTAL	2 TOTAL	1 WHITE	2 WHITE	1 NEGRO	2 NEGRO
1	Total	↑ .54	.25 ↓	.39	-.07	.35	.24
2		.23	.44	.24	.25	.10	.29
1	White	.44	.33	↑ .62	.15 ↓	.06	.08
2		.12	.38	.25	.39	.03	.10
1	Negro	.57	.17	.20	-.46	↑ .80	.34 ↓
2		.37	.63	.20	-.05	.37	.74

to be background and experience as measured by the dimensions of the Personal History Index. The variable Early Family Responsibility, which occurred across racial groups in predicting the paired-comparison performance rating, occurs again with a positive weight for the CPD rating but only in two sets of regressions for white groups. A new variable, Stability, has positive weights in both white and Negro regressions. The remaining predictors are race-specific, with positive weights shown by white groups for the variables Professional-Successful Parents, which also occurred for the paired-comparison criterion, and General Health, and with additional positive weights for the variables School Activities and Higher Educational Achievement. One Negro group shows a positive weight for Vocational Decisiveness, and the variable Financial Responsibility, which occurred previously with rather anomalous negative weights, again appears on two Negro groups but with inconsistent signs.

The Test of Social Insight also provides some useful predictors for the Chicago-Police-Department rating. The variable Cooperativeness appears again with a positive weight in three out of the four sets of regressions for the white group. Competitiveness again appears in a white regression with a negative weight. A new variable, Aggression, appears with a negative weight for a Negro group. Although Passivity occurs across racial groups, its signs are inconsistent for the Negro group but negative for the white group.

The behavioral measures derived from the Temperament Comparator appear to be more important for differential prediction for the racial groups. Eight variables emerge as predictors, but of these only two, the traits Prompt Starter with positive weights and Demonstrative with negative weights, predict in both white and Negro regressions. The remaining predictors with positive weights for the white group are the traits Self-Confident and Steady Worker and with negative weights, the trait Persevering and the factor Sociable. The Negro group shows a positive weight on the factor Controlled and a negative weight on the trait Socially at Ease.

The Arrow-Dot Test provides predictors only for the Negro group with a positive weight on Ego Control and a consistent negative weight on the Superego variable.

In the area of mental abilities, both the Closure Speed and Perceptual Speed tests are predictors for the white group but in a negative direction. The Closure Speed test does, however, predict positively for a Negro group. The results in the mental ability area will need to be analyzed further, since these are the ones most likely to be affected by the civil service screening of patrolmen.

Of the Wave-specific tests, the most useful appear to be the Work Interest Index, the Edwards Personal Preference Schedule,and the Cree Questionnaire.

PREDICTION OF TENURE

In attempting to predict tenure, we are predicting a rather different dimension of patrolman behavior than that represented in the paired-comparison and CPD performance ratings. One may thus expect rather different variables to appear as predictors. Another way in which the tenure criterion differs from the performance ratings is in the obtained validity coefficients. Although the within-race coefficients are still greater than those obtained from racially-mixed groups, the disparity between them is much less. Therefore, one may expect a greater interchangeability in the test variables which predict this criterion for the two racial groups.

It will be seen from Tables 6 and 7 that the Personal History Index again produces a sizable number of predictors, seven. Of these seven, four have consistent weightings across racial group regressions. These are positive weights for Stability, Financial Responsibility, and Higher Educational Achievement, and negative weights for Vocational Satisfaction. In addition, the white groups show positive weights on Early Family Responsibility and Parental Family Adjustment and a negative weight on Leadership & Group Participation. These results are in accordance with those obtained in predicting tenure for other

CHAPTER VI—Table 6

Validation and Cross-Validation Results for Tenure within Wave I

MULTIPLE CORRELATION	PART	1		2		1		2		1		2	
	GROUP	TOTAL		TOTAL		WHITE		WHITE		NEGRO		NEGRO	
	N		80		101		58		73		22		28
	m		11		12		7		14		6		8
	R		.69		.68		.70		.66		.80		.94
	R′		.62		.62		.65		.55		.70		.92
PREDICTORS	var #	r	β	r	β	r	β	r	β	r	β	r	β
PHI 2 Higher Educational Achievement	2	.28	.17	.39	.33	.29	.14	.39	.31	.32	.08	.39	.40
5 Financial Responsibility	5	.30	.22	.31	.15	.30	.24	.27	.13	.44	.16	.45	.23
6 Early Family Responsibility	6			.25	.11			.26	.21				
8 Stability	8	.42	.10	.56	.21	.40	.06	.47	.02	.70	.50	.85	.51
13 Vocational Satisfaction	13	-.23	-.11	-.40	-.18			-.34	-.09			-.59	-.24
NON-VERBAL REASONING	16*			-.18	.11								
UNDERSTANDING COMM.	17*			-.26	-.04			-.30	.14				
CLOSURE SPEED	18			-.26	-.06			-.26	.01				
PERCEPTUAL SPEED	20			-.27	.05			-.34	-.02				
AC III Originality	24*	-.23	-.16										
PRESS I	25			-.30	.02			-.29	-.15				
II	26	-.22	-.03	-.33	-.07			-.30	.27				
III	27	-.42	-.14	-.38	-.16	-.46	-.31	-.36	-.30			-.43	-.10
II-III	29	.29	.24			.31	.17						
TC I Controlled	31							.23	-.06				
b Cautious	37											-.37	-.07
d Demonstrative	39							-.24	.16				
k Prompt Starter	46	.26	.19			.28	.24			.30	.15		
HAND Environmental	71*									-.30	-.22		
MPI Extroversion	84*							-.21	-.21				
A-D I Score	96	.29	-.04									.38	.08
E Score	97	-.32	-.26			-.33	-.28			-.39	-.21	-.46	.08

*Unique to Wave I

CROSS-VALIDATION	PART	GROUP						
	1	Total	.69	.52	.67	.47	.53	.50
	2		.57	.68	.51	.66	.57	.65
	1	White	.72	.52	.70	.47	.51	.49
	2		.50	.62	.45	.66	.49	.57
	1	Negro	.70	.62	.62	.48	.80	.70
	2		.77	.89	.69	.62	.83	.94

CHAPTER VI—Table 7

Validation and Cross-Validation Results for Tenure within Wave II

MULTIPLE CORRELATION	PART GROUP	1 TOTAL		2 TOTAL		1 WHITE		2 WHITE		1 NEGRO		2 NEGRO	
	N		97		95		61		65		36		30
	m		22		17		19		12		8		9
	R		.85		.77		.89		.83		.86		.81
	R′		.80		.70		.84		.78		.81		.71
PREDICTORS	var #	r	β	r	β	r	β	r	β	r	β	r	β
PHI 2 Higher Educational Achievement	2	.20	.07	.23	.06					.35	.26	.40	.23
4 Leadership and Group Participation	4	-.33	-.17			-.35	-.19						
5 Financial Responsibility	5	.49	.09	.40	.08	.59	.09	.41	.05			.46	.23
6 Early Family Responsibility	6	.21	.05	.20	.07			.26	.10				
7 Parental Family Adjustment	7							.26	.15				
8 Stability	8	.76	.41	.62	.20	.80	.47	.63	.29	.69	.43	.68	.11
9 School Activities	9	-.20	.13										
13 Vocational Satisfaction	13	-.63	-.30	-.56	-.38	-.65	-.17	-.56	-.34	-.62	-.33	-.54	-.18
15 General Health	15			.26	.09								
CLOSURE SPEED	18	-.30	-.14	-.22	-.09	-.33	-.05						
CLOSURE FLEXIBILITY	19	-.28	.05			-.32	-.00						
PERCEPTUAL SPEED	20	-.27	.05							-.35	.25		
PRESS I	25	-.36	-.08	-.33	.02	-.37	-.14	-.41	-.11	-.37	.16		
II	26	-.38	-.10	-.38	-.14	-.34	.17	-.42	.16	-.48	-.46		
III	27	-.34	.07	-.31	-.07	-.35	-.17	-.26	-.32	-.33	.02	-.40	-.22
II-III	29							-.27	-.17				
TC D Consistency	30	.27	.10										
V Sociable	35					-.27	-.11						
b Cautious	37											-.36	-.21
g Steady Worker	52			.25	.11			.29	.14				
A-D I Score	96	.29	-.06	.29	.30	.39	-.09					.40	.20
E Score	97	-.30	-.06	-.33	.42	-.40	-.20	-.24	-.06			-.45	.18
S Score	98			.22	.27								
WII 5 Artistic and Stylized	645*					.33	-.05						
8 Clerical and Routine	648*	.23	-.01			.30	.08						
Flexibility	653*					.27	-.11						
Aspiration Level	654*	.21	-.00			.34	.05						
CREE 4 Unstructured vs. Unstructured Situation	660*			-.27	-.09			-.39	-.17				
12 High vs. Low Artistic	668*	-.20	-.09										
EPPS Deference	671*	.22	-.02			.30	.02						
Order	672*			.27	.07								
Dominance	678*	-.21	-.05			-.31	-.08						
Change	681*	-.28	-.06							-.36	-.19	-.37	-.33
Endurance	682*					.26	.07						
Heterosexuality	683*			-.24	-.02								

*Unique to Wave II

CROSS-VALIDATION PART	GROUP	1 TOTAL	2 TOTAL	1 WHITE	2 WHITE	1 NEGRO	2 NEGRO
1	Total	.85	.75	.80	.75	.77	.66
2		.69	.77	.66	.74	.68	.57
1	White	.88	.77	.89	.79	.73	.63
2		.69	.80	.67	.83	.67	.44
1	Negro	.81	.71	.63	.68	.86	.71
2		.71	.70	.69	.59	.71	.81

occupational groups. However, a word of caution must be introduced here. The use of the variable Stability for the prediction of tenure for currently employed personnel will cause a spurious increase in the magnitude of the obtained coefficients. This variable was designed and intended for use in predictive validity studies.

The Press Test also produces some predictors for this criterion. Part III of the test is consistent in its negative weightings in seven out of the eight possible white and Negro regressions, and Part I may aid in the white regressions. The fairly frequent changes in sign between the zero order correlations and their corresponding beta-weights suggest that Part II may be acting as a suppressor variable in this system of test variables.

The Temperament Comparator is of less importance in the prediction of tenure, and only a few scattered variables appear. On the Arrow-Dot Test, there are some negative weights on the Ego Control variable for the white group but the Negro group shows fluctuations in the beta-weight signs.

Of the Wave-specific tests, the Edwards Personal Preference Schedule produces one variable each for the white and Negro groups and the Cree Questionnaire a variable for the white group. There are a number of variables from the Work Interest Index, but they have low beta-weights in this concurrent validation dealing with currently employed patrolmen.

IDENTIFICATION OF THE IMPORTANT PREDICTOR VARIABLES IN THE STUDY

The important predictor variables which have emerged from the eight sets of multiple-regression analyses, performed for each of the three major criterion measures in the study, cover all the areas of measurement included in the test batteries.

Motivational Measures

The motivational measures of background and experience, derived from the Personal History Index, are more important than those of work interests. In fact, the Personal History Index variables contribute more to the prediction of all three criteria than do any others in the study. However, work interests, as measured by the Work Interest Index, cannot be ignored since they do make some contributions to all three criteria. Although their beta-weights are low for the prediction of tenure, one could reasonably expect that these work-interest variables would increase in importance in the selection of new recruits, especially those who have not had much prior work experience to be assessed by the Personal History Index.

Intellectual Measures

The intellectual measures include both specific mental abilities and special aptitudes. Of the specific mental abilities, those dealing with visual perception appear to the most important. However, for groups of white patrolmen the beta-weights on Closure Speed and Perceptual Speed are negative. Negro patrolmen show conflicting positive and negative weights on Closure Speed. Actually, the whole mental abilities area is the one most likely to have been affected by the civil service screening of patrolmen, which eliminates candidates with IQs of less than 100. Thus, in testing currently employed patrolmen, we are dealing only with a curtailed sample with respect to intelligence. Final conclusions concerning the required mental abilities for success as a patrolman must await a predictive validation study based on unscreened applicants to the police force.

Of the special aptitudes, those dealing with social insight, as measured by the Test of Social Insight, are the most important. In fact, the Test of Social Insight ranks second only to the Personal History Index in the prediction of the performance criterion measures, although it does not contribute to the prediction of tenure. The Cree Questionnaire, a measure of creative potential, is another useful aptitude test for the prediction of performance. However, its inclusion in a battery would depend upon the other tests selected and the time available for testing.

Behavioral Measures

The most useful test in this third area is the Temperament Comparator, which ranks with the Test of Social Insight in the prediction of the performance criteria. It also has scattered variables which make some contribution to the prediction of tenure. Of the remaining tests in the behavioral area, those which contribute somewhat to all three criteria are the Press Test, the Arrow-Dot Test, and the Edwards Personal Preference Schedule. The Press Test and the Arrow-Dot Test have time limits of five and six minutes respectively. The Edwards Personal Preference Schedule is untimed, and patrolmen took between 30 and 40 minutes to complete it. Thus, the Press Test and the Arrow-Dot Test make greater contributions relative to expended testing time and are therefore to be preferred in the battery.

Some of the important areas of measurement and the specific variables contributing to them have been identified by the multiple-regression analyses. The "mortality rate" of tests selected for try-out in the batteries was gratifyingly low, indicating an initially relevant selection of tests for the Wave I sample. Furthermore, tests which were substituted in the Wave II sample were generally more predictive than those dropped from the Wave I sample. Hopefully, this is a result of the experience gained from the initial validations. We should be in a very good position to select tests for a predictive validation study.

PATTERNS OF TEST-VARIABLE SCORES LIKELY TO BE MOST PREDICTIVE OF PERFORMANCE

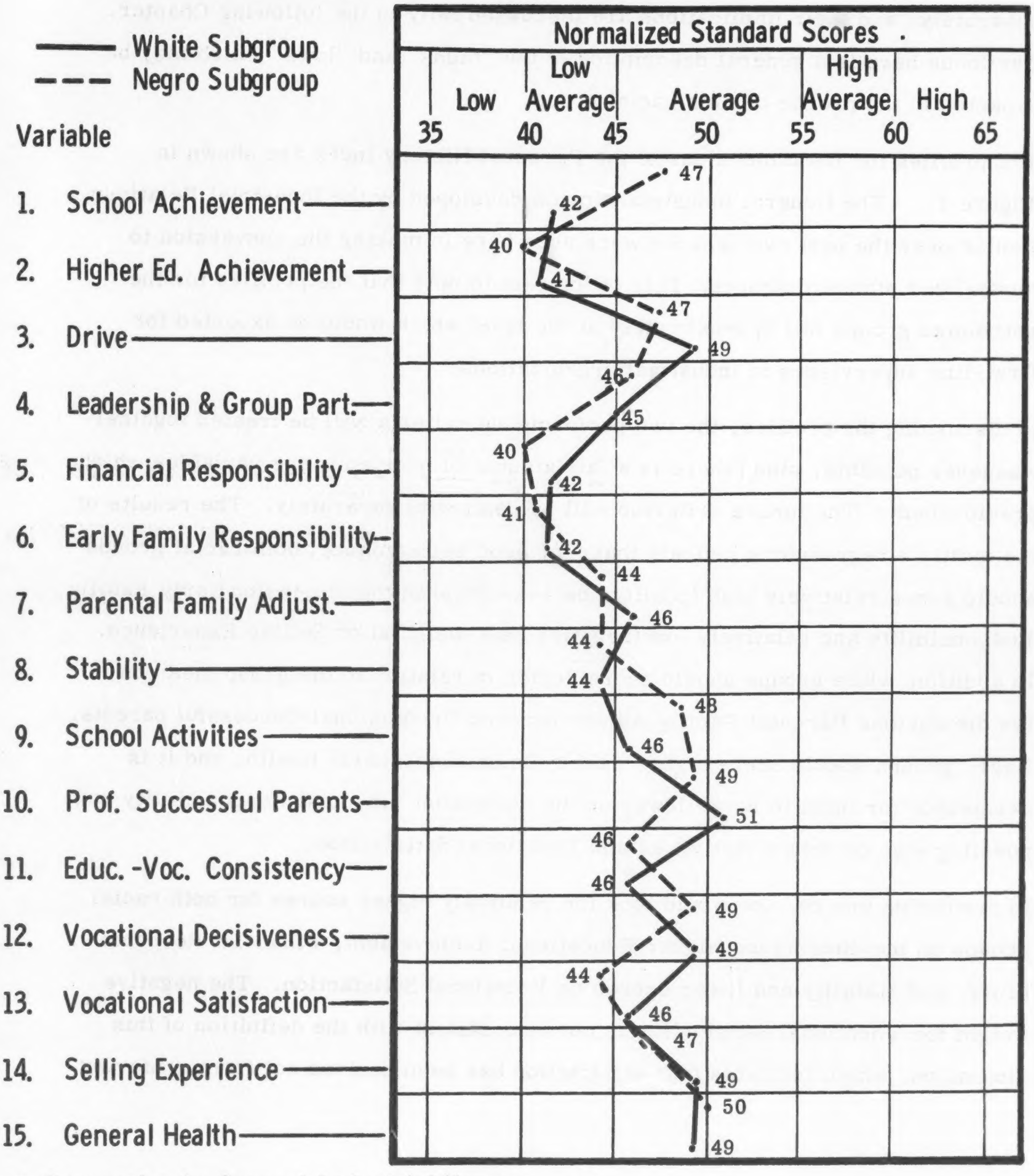
Our next aim in this Chapter is to describe the patterns of test-variable scores likely to be most predictive of good performance on the part of currently employed patrolmen. The first step in arriving at these patterns was to combine each racial group from Wave I with the similar group from Wave II. Then the respective mean scores for the test variables were calculated and converted to a normalized standard score scale as described in the Appendix of Chapter III. In order to keep the sample constant, only tests which were common to both Wave samples were used. The converted scores were profiled separately for the white and Negro groups. An examination of the profiles will show that there are some differences between the white and Negro mean scores. Significant racial differences, not only for the total sample but for the Wave I and Wave II samples separately, and their implications are discussed fully in the following Chapter. Our focus here is a general description of the "highs" and "lows" which may be expected in the profile of each racial group.

The profiles for the dimensions of the Personal History Index are shown in Figure 1. The General Industrial Norms developed by the Industrial Relations Center over the past two decades were used here in making the conversion to normalized standard scores. It is of interest to note that the profiles for the patrolman groups fall approximately at the level which would be expected for first-line supervisors in industrial organizations.

In discussing the profiles, the two performance criteria will be treated together whenever possible, since there is a fair amount of overlap in the variables which predict them. The tenure criterion will be dealt with separately. The results of the multiple regressions indicate that, for good performance, both racial groups should score relatively high (positive beta-weights) on the dimension Early Family Responsibility and relatively low (negative beta-weights) on Selling Experience. In addition, white groups should score higher in relation to the group mean on the dimensions Parental Family Adjustment and Professional-Successful parents. Negro groups should score higher on the dimension General Health, and it is acceptable for them to score lower on the dimension Financial Responsibility and possibly also on School Activities and Vocational Satisfaction.

In predicting tenure, one would look for relatively higher scores for both racial groups on the dimensions Higher Educational Achievement, Financial Responsibility, and Stability and lower scores on Vocational Satisfaction. The negative weight for Vocational Satisfaction is not inconsistent with the definition of this dimension, which indicates that satisfaction has been derived not from achieving

CHAPTER VI -- Figure 1
Profiles Based on Mean Scores*
of the Combined White (N=358) and Combined Negro (N=154) Subgroups
on Dimensions of the Personal History Index



* Conversion by General Industrial Norms

high goals but rather from an easy acceptance of the occupational requirements to be met. In addition, white groups should show higher scores on Early Family Responsibility and Parental Family Adjustment but lower scores on Leadership & Group Participation.

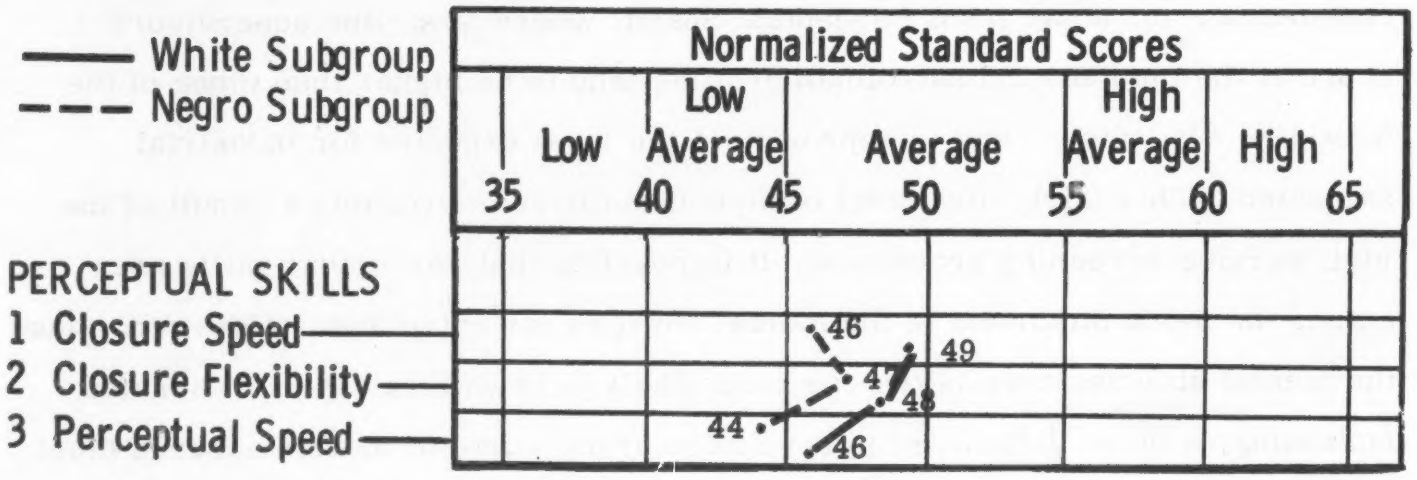
The profiles on the three mental ability tests common to both Wave samples are shown in Figure 2. Here again the General Industrial Norms were used for conversion, and apart from Perceptual Speed, where first-line supervisors score at the median, the patrolman profiles tend to be higher than those of the first-line supervisor and to approximate the level expected for industrial salesmen. This fairly high level of mental abilities is probably a result of the civil service screening procedure. It is possible that perceptual skills are among the more important of the mental abilities for patrolmen. However, since the mental abilities area is the one most likely to be affected by civil service screening, a more definitive interpretation of the worth of these measures must await a predictive validation study.

The profiles for the Test of Social Insight are shown in Figure 3. Since this test has not been previously used by the Industrial Relations Center, the conversion was made using the Adult Male Norms supplied in the test manual. Thus a comparison of patrolmen with the general industrial population is not possible. The results of the multiple-regression analyses indicate that for successful performance both racial groups should score higher on Cooperativeness and lower on Withdrawal and Competitiveness in social situations. In predicting the CPD performance rating, there are indications that a low score on Passivity would be desirable.

The profiles for the factor scores and the trait scores from the Temperament Comparator are shown in Figures 4 and 5. The first observation of importance is that although there are some differences in level between the profiles for the white and Negro subgroups, there is a perceptible similarity in the shapes of the two profiles. Since both groups are employed patrolmen, this finding in itself attests to the stability or reality of this occupational profile. The General Industrial

CHAPTER VI -- Figure 2

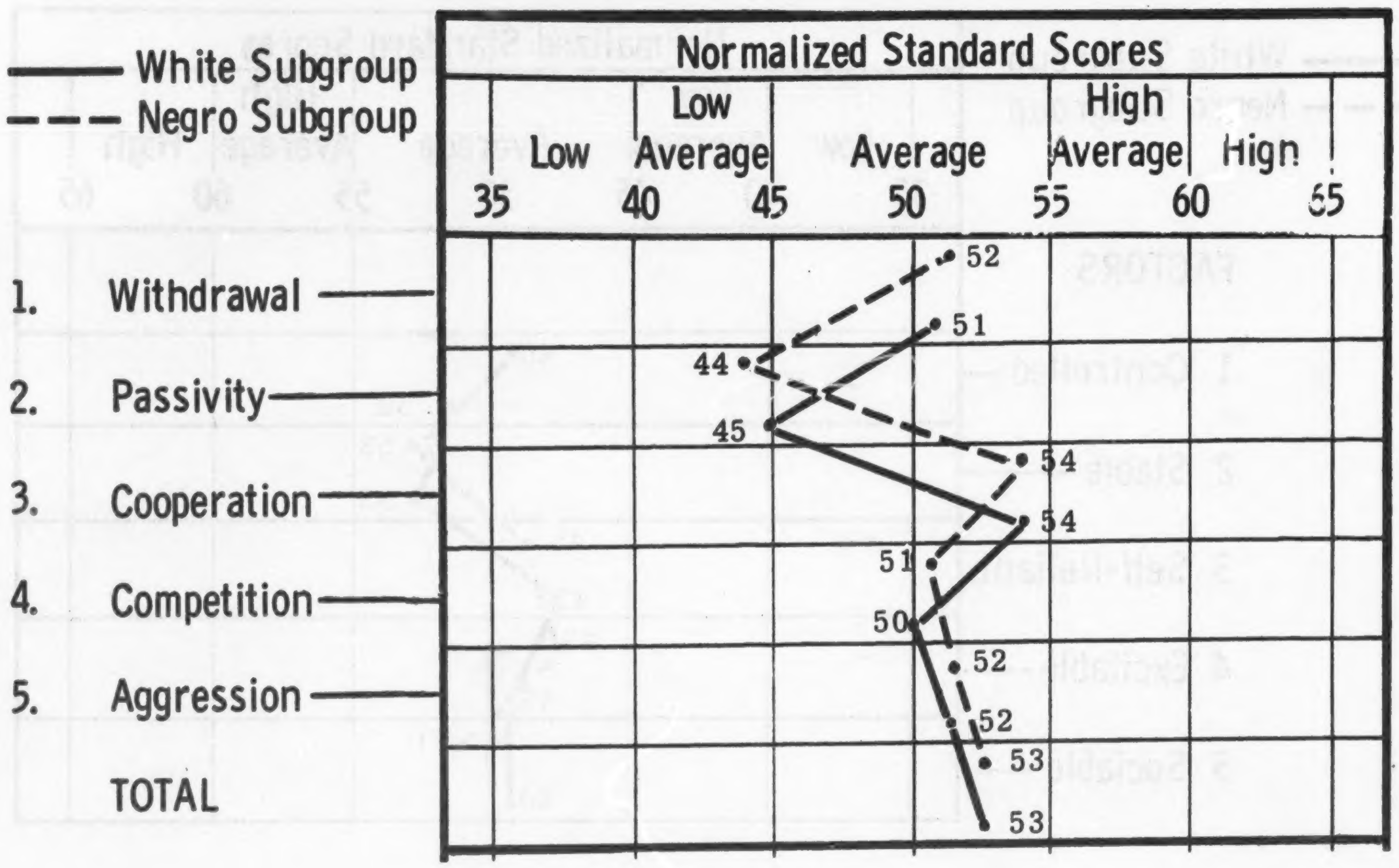
Profiles Based on Mean Scores*
of the Combined White (N=358, 355, 346) and
Combined Negro (N=154, 153, 151) Subgroups
on Tests of Specific Mental Abilities



* Conversion by General Industrial Norms

CHAPTER VI -- Figure 3

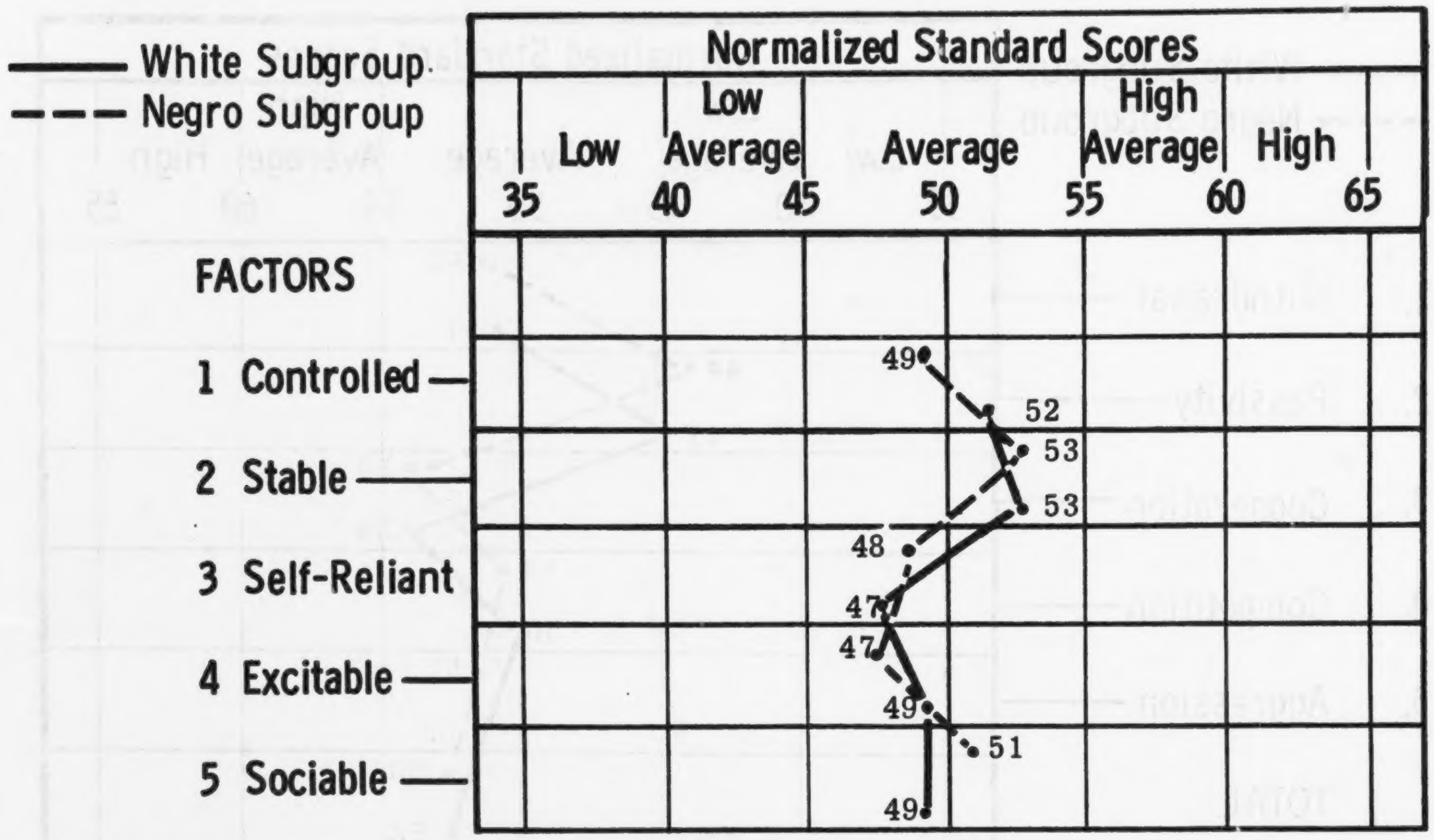
Profiles Based on Mean Scores*
of the Combined White (N=355) and Combined Negro (N=152) Subgroups
on Dimensions of the Test of Social Insight



* Conversion by Adult Male Norms supplied in the test administration manual

CHAPTER VI -- Figure 4

Profiles Based on Mean Scores*
of the Combined White (N=330) and Combined Negro (N=139) Subgroups
on the Factor Scores from the Temperament Comparator



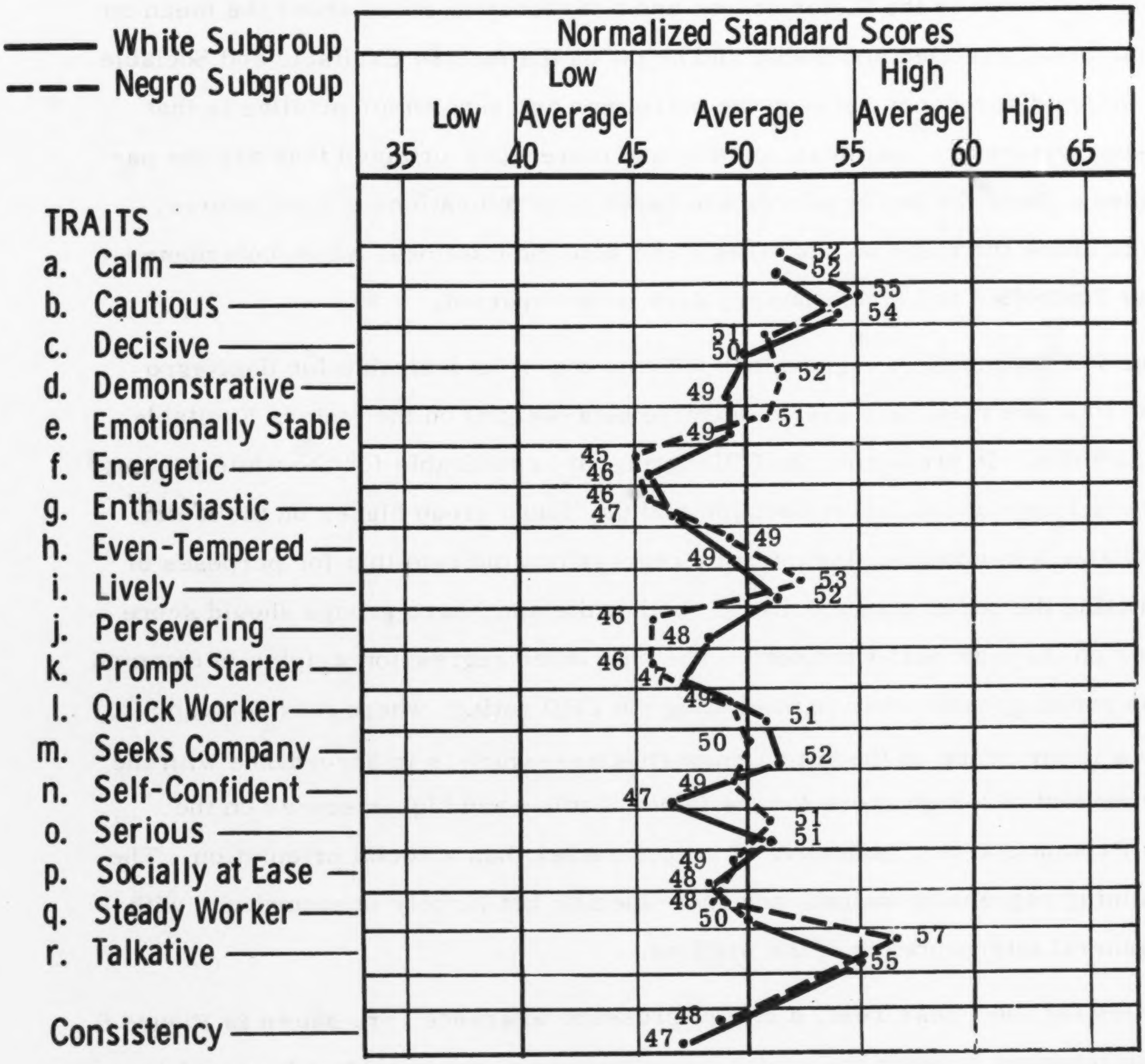
---

* Conversion by General Industrial Norms

3

CHAPTER VI -- Figure 5

Profiles based on Mean Scores* of the Combined White (N=330) and Combined Negro (N=139) Subgroups on the Trait Scores from the Temperament Comparator



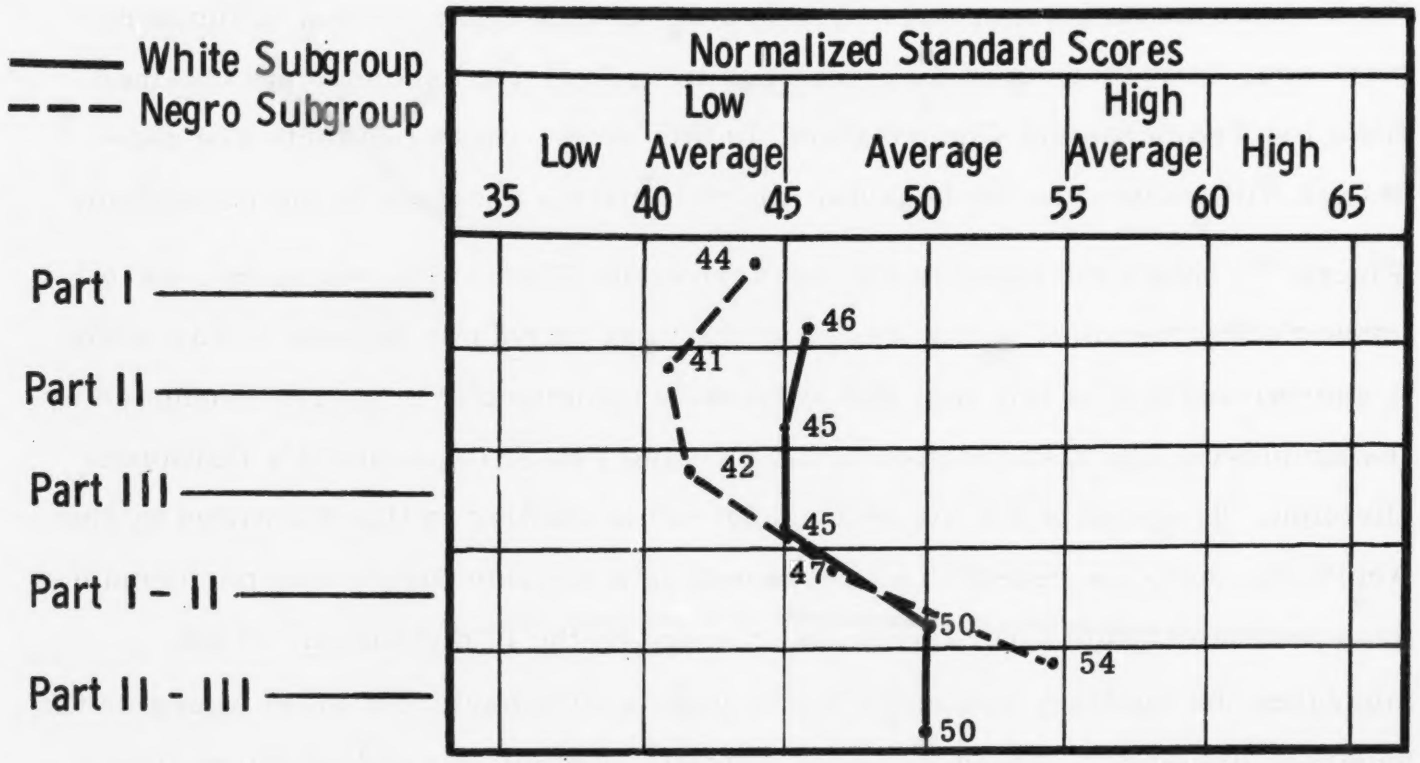
* Conversion by General Industrial Norms

Norms were used for conversion, and again the patrolman samples are closer to the scores of the first-line supervisor than they are to those of other industrial groups, such as hourly, sales, or executives. The general characteristic of the patrolman profiles, like that of the first-line supervisors, is a lack of extreme variations in the factor scores and a tendency to score above the mean on the factors Controlled and Stable and below on the factors Excitable and Sociable. The major discrepancy between the patrolman and supervisor profiles is that the supervisors are even less socially and more work oriented than are the patrolmen. Since the factor scores are based on combinations of trait scores, scores above the mean on such traits as Calm and Cautious, which contribute to the Controlled and Stable factors, are to be expected.

Results of the multiple regressions indicate that it is desirable for the Negro group to score relatively lower (negative beta-weight) on the factors Excitable and Sociable. In predicting the CPD rating, it is desirable for the white group to score lower on the factor Sociable and the Negro group higher on the factor Controlled. In addition, the multiple regressions indicate that for purposes of predicting the performance criteria, both white and Negro groups should score higher on the trait Self-Confident. The only other regression weights in common to the racial groups occur in predicting the CPD rating, where it is desirable to show a lower score on the trait Demonstrative--which is in accordance with the requirement of a high score for the factor Stable--and higher scores on the trait Prompt Starter, indicative of a work rather than a social orientation. The remaining regression weights are race-specific but largely in accordance with the general interpretation of the profiles.

Profiles for the Press Test, a test of pressure tolerance, are shown in Figure 6. The differences between the racial group means are more evident here and are discussed in detail in the following Chapter. However, once again the general shape of the profiles for the two racial groups is similar. There is some evidence from the regression analyses that Parts I and III of the test may be

CHAPTER VI -- Figure 6
Profiles Based on Mean Scores*
of the Combined White (N=353) and Combined Negro (N=154) Subgroups
on Dimensions of a Test of Pressure Tolerance



* Conversion by General Industrial Norms

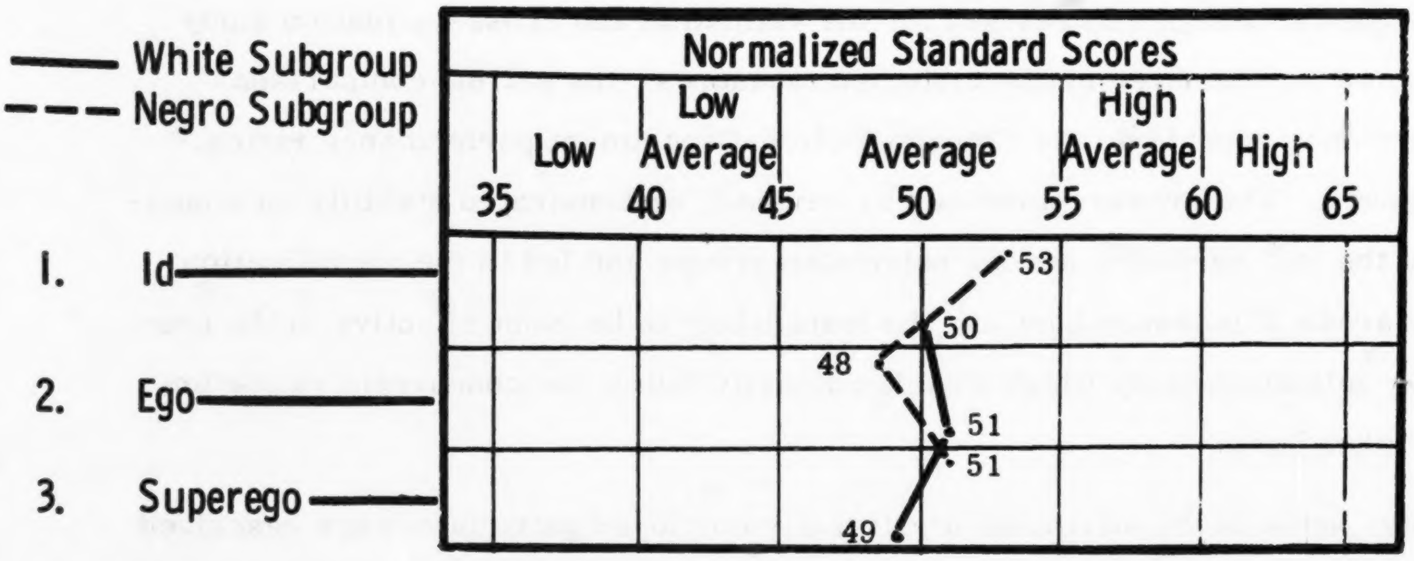
predictive of tenure for both racial groups. There is also evidence that the derived scores may be predictive of the CPD rating for the white subgroup. With respect to the interpretation of the scores, it should be pointed out that scores below the median for the first three parts of the Press Test have been found to be associated with high scores on the Controlled and Stable factors of the Temperament Comparator. The results of this objective test of temperament are, therefore, used as a check on the subjective assessments obtained from the Temperament Comparator. In this study, these two tests are consistent with respect to the behavior characteristics ascribed to the patrolmen.

Figure 7 shows the profiles for the Arrow-Dot Test. The raw scores were e converted by means of norms based on the total patrolman sample in this study. A shorter version of this test had previously been used in research conducted by the Employee Appraisal Section of the Chicago Police Department's Personnel Division. The profile for the white subgroup is similar to that identified by the Appraisal Section's research as representing successful patrolman performance. This profile exhibits a relatively lower score on the Id dimension, which measures the tendency to uncontrolled, impulsive behavior, a lower score on the Superego dimension, which measures interior controls viewed as external restrictions, and a higher score on the Ego dimension, which is essentially a measure of the ability to function realistically. The results of the regression analyses accord with these interpretations. Thus, in predicting the paired-comparison performance rating, it is desirable for the white group to score low on the Id dimension, and in predicting the CPD rating, it is desirable for the Negro group to score low on the Superego dimension and high on the Ego dimension.

SUMMARY

The basic objective of the validation of the occupational test batteries was to determine the extent to which the selected tests could account for differences in the performance of patrolmen. The results of the preliminary regression analysis indicated significant and acceptably high relationships between the tests and all eight of the performance criterion measures used in the study. These

CHAPTER VI -- Figure 7
Profiles Based on Mean Scores*
of the Combined White (N=358) and Combined Negro (N=154) Subgroups
on Dimensions of the Arrow-Dot Test



* Conversion by Norms Based on the Total Patrolman Sample

relationships were increased slightly by the substitution of some additional tests in the battery used for the second sample of patrolmen.

The preliminary analysis also revealed that higher validity coefficients (or more significant relationships) were obtained for the racial groups treated separately than for the racially-mixed total group. As a result of this finding, the original experimental design was revised for the validation and cross-validation study conducted for the three major criterion measures--the paired-comparison performance appraisal, the Chicago-Police-Department performance rating, and tenure. The cross-validation, as revised, demonstrated stability on a number of the test variables across patrolman groups and led to the identification of the areas of measurement and the tests likely to be most effective in the predictive validation study which would ordinarily follow the concurrent validation undertaken here.

Finally, some of the attributes of currently employed patrolmen were described through the interpretation of group profiles on the predictive tests. A consistent picture emerged of the personal background characteristics, some of the special skills required in social situations, and of the behavior attributes which were likely to contribute to the successful performance of police officers.

REFERENCES

Equal Employment Opportunity Commission. Guidelines on employment testing procedures. Washington, D.C.: EEOC, August 24, 1966.

Guilford, J.P. Fundamental statistics in psychology and education. (3rd ed.) New York: McGraw-Hill Book Company, Inc., 1956.

CHAPTER VII

Validation of an Occupational Test Battery for a Racially-Mixed Group

The total occupational group of beat patrolmen in the Chicago Police Department comprises two major racial subgroups. At the present time, about three quarters of its membership is drawn from the white population and the remaining quarter from the Negro. Early in the study, it became clear that the racially-mixed composition of the group would require the use of special procedures and techniques, particularly during the validation of the occupational test battery. The problems thus introduced were not unique to this project. They are inseparable from larger issues and controversies which have only recently risen to a prominent place in psychology and in society at large. Below is a brief historical review of the major developments--all within the last decade--which supply the framework for the validation section of the Chicago Police Department study.

HISTORICAL REVIEW

Two primary lines of events will be traced in this review. The first is an attack upon psychologically-designed methods for the selection of personnel, including, but not confined to, the use of psychological tests. The second is the passage of the Civil Rights Act of 1964, particularly its Title VII, which deals with fair employment practices and equal employment opportunity.

The attacks upon psychologically-based personnel selection methods began slowly but grew to storm proportions in a series of much-publicized and bitter criticisms of testing written by such laymen as William H. Whyte, Jr. (1956), Martin L. Gross (1962), and Hillel Black (1963). The substance of these criticisms was varied--tests discriminated against the creative, or against the intelligent, or against the culturally deprived; test questions were offensive or intrusive; test interpretations were inaccurate or arbitrary or otherwise prejudicial to the persons tested. Eventually, the whole issue came to the attention of the Congress, and the controversy was climaxed in June, 1965, by a full-scale Congressional investigation, involving committees of both the Senate and the House, of the use of psychological tests. The two major bases for the legislators' criticisms of testing stemmed from two political orientations. Conservatives and right-wingers concentrated on the issue of invasion of privacy, while liberals and leftists attacked tests because they might discriminate against disadvantaged groups in the labor market.

It is the question of discrimination which concerns us here. The 1964 Civil Rights Act expressly prohibits discrimination of many different forms and in a variety of contexts. Title VII of the Act focuses on the context of employment. Its provisions forbid employers, labor unions, and employment agencies to discriminate on the basis of race, color, religion, sex, or national origin, except as specified in Section 703 (e), that is, except in instances where religion, sex, or national origin can be demonstrated to be a bona fide occupational qualification.

Of particular interest is a clause in Section 703 (h) of Title VII, known as the Tower amendment. This amendment was designed to protect the right of management to use tests for purposes of employment, placement, and promotion. It was directly related to the renowned and then current case of Myart vs. Motorola, which was heard in 1963 by the Illinois Fair Employment Practices Commission.*
In the amendment's final form, it reads essentially as follows:

> Notwithstanding any other provision of this title, it shall not be...
> an unlawful employment practice for an employer to give and to

---

*The Commission's findings in favor of Myart were later reversed by a March 24, 1966 decision of the Illinois Supreme Court.

> act upon the results of any professionally developed ability test, provided that such test, its administration or action upon the results is not designed, intended or used to discriminate because of race, color, religion, sex or national origin.

In short, employment tests may be used, but they must not be intended to discriminate against any subgroup within the population. They must be used properly and professionally. As a result of the Tower amendment, the proper and professional use of tests has become a subject for much hard thought and much careful research. Two principal interested parties have been responsible for the extensive literature on this topic which has been developing steadily since 1964. First, of course, are the professional psychologists in academic and organizational settings, the people responsible for developing and using psychological tests. Second are the groups and individuals charged with the actual administration and enforcement of Title VII, most importantly the Equal Employment Opportunity Commission.

This Commission, established by Title VII and appointed by the President, began its operation shortly after the Act was passed. A task of immediate importance was to assist employers in implementing both the letter and the spirit of the new laws. In this project, the hotly-argued issue of testing was crucial, and the Commission set out to establish sound and practical guidelines for using tests in employment procedures. During the formulation of these guidelines, the Commission made a point of consulting closely with a panel of professional psychologists and drawing heavily on their thinking and research. For example, the March 1966 research report by Wallace, Kissinger, and Reynolds (1966), Testing of Minority Group Applicants for Employment, drew on current studies by Krug, Ash, Campbell, Short, and Brayfield.

In the more popularly oriented EEOC Guidelines on Employment Testing Procedures (1966), the Commission based its specific recommendations for the use of testing on the points advocated by this panel and by other consultants (Baxter, Docter, Elias, and Fein). The suggestions in this Guidelines report will be discussed below in the course of a review of various types of discrimination which can arise during employment testing. Two broad classes of discrimination will be described. The first is blanket discrimination, which can affect anyone. The second is specific discrimination, which can operate against particular subgroups within the total population. The 1965 Congressional investigation of testing called attention to both these classes. Title VII of the 1964 Civil Rights Act is directed at the latter.

BLANKET DISCRIMINATION

Blanket discrimination can take a number of forms, and may be either intentional or unintentional.

Deliberate Discrimination

It is obvious that test results can be used to discriminate against any individual or group if they are deliberately falsified or misreported. Standards for passing tests can also be deliberately shifted and set impossibly high for "unwanted" individuals or groups. However, tests are not unique as tools of discrimination. Other aspects of the selection process, such as the personal interview or the type of educational qualifications demanded, can be exploited as easily and probably more "safely." Such practices are of course illegal and fortunately not only fairly readily detectable but also, in practice, rare. Of far greater concern is discrimination resulting from carelessness or ignorance.

Improper Administration of Tests

Test scores are useful and valid only when they are obtained under standard conditions. For example, slipshod administration of tests may lead to variations in time limits for timed tests, which would plainly discriminate against those applicants who performed under the shorter time limits. Similarly, any variations in test instructions or in the handling of questions about such instructions can result in scores which are not comparable from one applicant or group of applicants to another.

The unintentional discrimination which can result from such practices is easily eliminated by careful training of all personnel involved in the selection procedure. Adequate professional training is particularly important for everyone responsible for actual interpretation of tests or for the overall administration of a selection and placement program. The necessity for professionalism has been a watchword of psychologists for many years, and national, regional, and state psychological associations both set and enforce requirements as to the qualifications, training, and ethical standards of their members.

The EEOC Guidelines particularly stress the need for qualified personnel. "...it is essential that tests be administered by personnel who are skilled not only in technical details, but also in establishing proper conditions for test taking." (1966, p. 4.) They further suggest that if, in spite of all precautions, there is any possibility that a failing applicant was overly anxious about the test situation or misunderstood the instructions in some way, he should be retested.

Inappropriate Use of Tests

There are two senses in which the use of a test may be inappropriate. First, the test itself may be technically unsound as a measuring instrument, yielding faulty or unreliable measures. For judging the appropriateness of a test in this sense, the EEOC Guidelines suggest using the "Standards for Educational and Psychological Tests and Manuals," published jointly by the American Psychological Association, the American Educational Research Association, and The National Council on Measurement in Education (1966).

Second, a test may be technically sound and yield consistent measures, but be used for a purpose for which it was not designed or intended. For example, there are in existence a number of excellent tests of mechanical aptitude, yet it would probably be inappropriate to include one of them in a test battery for the selection of female clerical employees. However, an analogous situation could and does exist where applicants for semi-skilled or skilled mechanical jobs are required to pass tests--such as reasoning tests--which call for high levels of verbal comprehension, verbal fluency, or other linguistic skills. It is unlikely that results on such tests reveal the man's actual competence or incompetence as a mechanic, just as it is unlikely that a test of mechanical aptitude would reveal a stenographer's competence or incompetence in taking dictation.

Such inappropriate use of tests can be avoided by directly investigating the relationship between scores on a test and independent measures of performance on the job. This investigation gives information on the validity of the test for that occupation. In the EEOC Guidelines, the Commission encourages the use of "job-related ability tests" and recommends that tests be judged in relation to actual job performance rather than in relation to whatever abstract or concrete attribute they were designed to measure. Professional psychologists have long advocated that tests be validated specifically for the occupational groups with which they will be used.

Nevertheless, even if an employer observes all of the safeguards outlined above, he may still, with the best of intentions, inadvertantly discriminate against specific subgroups in the population, such as women or the culturally deprived. Since the latter are predominantly Negroes or foreign language groups, racial and/or national discrimination would be occurring. Indeed, such inadvertant discrimination will occur unless certain recently developed, specialized techniques are used.

DISCRIMINATION AGAINST RACIAL SUBGROUPS AND TECHNIQUES TO PREVENT IT

Theoretically, for a test to be racially non-discriminatory in its selection of applicants for any particular occupation, a given test score should have the same meaning in relation to performance for candidates of all races. In practice, it has been empirically demonstrated in recent years that different racial groups may score differently both on the test (predictor) measures and on the performance (criterion) measures. Note, however, that if a criterion measure actually reflects performance on the job (possible bias in criterion measures will be discussed later) and if a particular racial group scores generally lower both on the test and on the criterion measure, then the test cannot be held to discriminate against that race. But racial discrimination would occur when different racial groups score differently on the tests and these differences are not reflected in actual performance.

It is now generally accepted that Negro subgroups will score lower than white subgroups on a wide variety of ability tests--Dreger and Miller (1960), Dugan (1966), Lopez (1966), and Lucas (1953). The situation is summarized by Krug (1966, pp. 24-25) as follows:

> By usual psychometric standards, we can say that we "know" that mean scores for Negroes are lower than mean scores for whites on most tests of general ability, intelligence, academic aptitude, or whatever you choose to call these measures of "G." Score distributions on such tests are invariably overlapping, often to an appreciable degree, but means are significantly different in the typical study. The evidence of tests of specific aptitudes is far less complete, but the generalization can probably be supported that mean score differences favoring whites is the most common finding.

Ash (March, 1966, pp. 9-10) more specifically reports the results of a study done of jail inmates at the Chicago Cook County Jail. A sample of 106 whites and 100 Negroes was given the U.S.E.S. General Aptitude Test Battery.

> On each test of the GATB, the mean score for the "standard American adult population" is 100. For this sample, the means as follows: Numerical: N-70, W-89; Finger Dexterity: N-72, W-85; Motor Coordination: N-75, W-79; Manual Dexterity: N-75, W-84; General Intelligence: N-76, W-93; Form Perception: N-77, W-90; Verbal: N-84, W-92; Clerical Aptitude: N-85, W-95; Spatial: N-89, W-100. On every aptitude test but motor coordination, there was a large, significant mean difference in favor of the whites.

It is also widely accepted that the lower scores of Negro subgroups--like those of the white subgroup in the example above--are the result of social and cultural deprivation and are not necessarily caused by innate or genetic influence--

Campbell (1965), Klineberg (1963), Lockwood (1963, 1966), and Mayfield (1964). Some psychologists have tackled the problem of differential test response of different racial groups by using non-verbal tests which make less call on the language skills so obviously lacking in the educationally and culturally deprived.

However, there is some evidence that Negroes, as compared with the white population, do no better on non-verbal tests than they do on verbal ones (Bragg, 1967) and that non-verbal tests are not necessarily better predictors of performance for Negro groups (Kirkpatrick, 1967).

The attempt to use non-verbal tests to measure basic skills or abilities without the contaminating influence of educational or cultural deprivation was an early and not very successful approach to the problem of test discrimination against specific subgroups within the population. Other possibly more fruitful approaches which warrent further study are described below.

"Culture Free," "Culture Fair," and "Culture Equivalent" Tests

At this stage of the effort to develop procedures and techniques which will prevent test discrimination against specific population subgroups, the concept of a "culture free" test has been thoroughly repudiated. The task of constructing test items which will be equally novel and equally meaningful to people of all cultures is difficult to the point of impossibility. Even if such items could be produced, it is hard to see how performance on them could be meaningfully related to performance in specific cultural and occupational settings.

A more promising approach is the development of "culture fair" or "culture equivalent" tests. An excellent example of this approach is found in Schwarz who, for the past four years, has headed a West African project for the development of aptitude tests to be used for selection purposes. As described by Krug (1966), the "culture fair" test is designed for use with culturally disparate groups. Its items are based on stimuli or referents with which members of both cultures are equally familiar. Krug suggests that "culture common" might be a good term for such a test. With "cultural equivalent" tests, the approach is somewhat different. Here there is a different form of the test for each cultural group. Test items in each form are appropriate and meaningful for the group in question. A test of this sort for gauging the mechanical aptitude of West Africans used "cultural counterparts" for the tools, machines, and mechanical principles which would be familiar to the average American. The test proved to have good validity in predicting performance of West Africans in mechanical training programs.

However, it is probable that the chances of success in constructing "culture equivalent" tests would increase with the degree of disparity between the cultures. That is, constructing culture equivalent tests for West Africans and white Americans--where there is both ideational and geographical separation--might be simpler than constructing equivalent tests for Negro Americans and white Americans--where, in spite of differences, there is considerable cultural overlap. In view of this probability, a better approach might be to use some of the procedures discussed below.

The Role of Suppressor and Moderator Variables in Test Validation

The use of so-called "suppressor" and "moderator" variables becomes possible whenever we have a battery of several distinguishable tests at our disposal, but must use this battery as if it were a single test to make decisions concerning a single criterion. There will be an infinite number of ways to combine the scores on the separate tests into an overall composite index, and some of these ways will be more appropriate than others. The problem is to find the method of combination that maximizes the "cross-validity" of the test battery, and to demonstrate that this increment in cross-validity is large enough to be distinguished from a statistical illusion. "Cross-validity" refers to the correlation that will exist in a new sample between the composite index of potential performance based on the test battery and measures of actual job performance in the real work situation.

The most straightforward and often the safest way to proceed is simply to discard any test in the battery, or any item in a test, that cannot demonstrate cross-validity on its own merits. Total scores obtained by summing such items or tests cannot fail to work at least as well as the best single item or test entering the composite, and the EEOC Guidelines are founded on this fact.

Only slightly more complex, and still a safe way to proceed, is to combine the component items or tests using weights suggested by multiple-regression analysis. The weights derived from such analysis can compensate for any possible imbalance in the content of a test battery that is being used to predict a multifaceted criterion; abilities that are underrepresented in the test battery are multiplied by larger weights than those that are overrepresented, so that the sum of weighted scores will display a larger correlation in cross-validation.

This is a "safe" procedure so long as the indicated weights remain positive or merely approach zero. It is not uncommon, however, for one or more of the indicated weights to become negative, which leads to the paradox that we subtract something from the overall battery score because of good performance on some item or test. The item or test assigned the negative weight is known as a "suppressor variable," although its function is still that of compensating for imbalances in the content of the test battery. In particular, suppressor variables are useful in situations where Test A is correlated with some criterion and Test B is correlated with Test A but not with that criterion. Test B becomes a suppressor variable and serves the function of cancelling out contaminating influences in the raw scores on Test A. Thus, while suppressor variables are logically defensible, their use is valid primarily in the context of poor test construction and even then it is relatively unsafe unless unusually large samples have been studied to determine all the proper weights. If possible, it is preferable to invest this effort in better test construction.

So far, we have considered only additive techniques for combining the component test scores into a single composite. Multiplication provides another possible way of dovetailing the information in the test battery with the specifications of a criterion. The criterion itself may be of such a nature that there is more than one way to perform with complete success, whereas with an additive test battery there can be only one way to attain a maximum score. If multiplication of one score by another is allowed, this inequity can be corrected. In effect, this multiplication makes the score on one test--called a "moderator" variable--determine the weight that will be given to another test. The test whose weight is thus controlled may actually be used as a predictor for some individuals and as a suppressor for others. Using the moderator approach, the number of ways of obtaining the maximum composite score will be 2^m, where "m" is the number of "moderator" variables used.

If, in a given situation, there really is only one best way, or one possible way, to go about the job, the moderator approach can only lead to a bit of extra analysis; the methodology incorporates the means for its own invalidation when this is appropriate. On the other hand, in situations where the moderator is appropriate but untried, the results that will be obtained will at best be a compromise that does no one justice, and there will be nothing to flag this fact. Thus, if there is so much as the possibility, as there certainly is in respect to groups of interest to the EEOC, moderator methodology would seem to be a must. Examples in which the moderator contribution is significant have been presented by Kirkpatrick et al. (1967).

In one special case, the moderator methodology can be bypassed. This is the case in which the moderator "variable" is simply membership or non-membership in a group, e.g., an ethnic group. In this special case, the moderator methodology may be expected to yield results that are numerically indistinguishable from the results obtainable by studying the two groups separately from each other. However, even in this case, the moderator approach can lend a unity to the results that is otherwise lacking, and can be of benefit for using the total sample as a single sample for determination of the best weights.

Separate Multiple-Regression Validations for Different Racial Groups

In the report by Kirkpatrick et al. (1967, p. 94), a series of studies dealing with separate validations for different ethnic groups are summarized. One of the conclusions is that "The results clearly indicate the existence of differential patterns of validity for different ethnic groups. That is, the test scores may predict job performance for one ethnic group but not for others."

This conclusion is best illustrated by their description of a study of 125 culturally deprived and unemployed males belonging to three different ethnic groups (white, Negro, and Spanish). These men were being trained for heavy vehicle driving through the Basic Essential Skills Training program sponsored by the Port of New York Authority and the City of New York. For purposes of guidance and development, all subjects were given tests of numerical and of reading ability when they entered the program. However, all were accepted into the program regardless of their scores on these tests.

At the conclusion of the program, a proficiency test was given to serve as a measure of performance and as one criterion in the study. On this performance measure, there were no significant differences among the three ethnic groups. On the other hand, on the pre-program tests, the white group had scored significantly higher than either the Negro or the Spanish. Furthermore, calculation of multiple-regression validity coefficients showed the numerical test had significant validities for the white and Negro groups but not for the Spanish. In contrast, the test of reading ability showed significant validity only for the Spanish group. Clearly if these tests had been used to select participants for the program, they would not have been equally predictive for the three groups in question.

There is a further point of interest in the results of this study. Both of the pre-program tests showed significant validity for the total sample. This finding would have been extremely misleading if there had been no further step of calculating separate validities for each of the subgroups within the total sample.

In connection with this point, it should be noted that there is some conceptual confusion in the validation procedures suggested in the EEOC Guidelines statement that "The sample population (norms) used in validating the tests should include representative members of the minority groups to which the tests will be applied." (1966, p. 4) Representation of all subgroups in the norms will not, in fact, clarify the relationship between test scores and performance for the different ethnic groups involved. However, the intent of the Commission seems to be clear when they state, "Only a test which has been validated for minorities can be assumed to be free of inadvertant bias." (1966, p. 4) In view of the emerging consensus of opinion regarding the necessity for separate validations, the separate-validation approach was taken in analyzing the data obtained in the Chicago Police Patrolmen study.

RESULTS FOR CHICAGO POLICE PATROLMEN: MULTIPLE REGRESSION VALIDATIONS

Two independently selected samples of beat patrolmen, referred to as Wave I and Wave II, have been used throughout this study. In selecting these samples, every attempt was made to ensure that they were each representative of Chicago Police patrolmen with respect to:

1. Demographic and racial characteristics,
2. The varying operational conditions in different districts in the Department, and
3. The full range of performance on the job, as measured by the major performance criterion used in the study.

Details on the selection of the samples are given in Chapter III.

A discussion of the selection of the test battery administered to each sample and a definition of the test variables contributing to these batteries are given in Chapter IV. The tests were selected with a view to providing:

1. As wide coverage as possible of the essential skills in patrolman performance, previously identified in the occupational analysis,

2. As wide coverage as possible of general human behavior, skills, and traits, and
3. Objectivity and standardization of both administration and scoring procedures. (All tests used were of the "paper-and-pencil" variety, suitable for group administration, and amenable to objective scoring by nonprofessional personnel.)

A summary statement of the tests used for the two samples is given in Table 1.

The selection and definition of the eight criterion (performance) measures, which hopefully can be predicted by test battery results, are described in Chapter III. Chapter VI analyzes the predictor (test battery) variables which best predict these criterion measures and discusses the details of the primary and cross validations for both samples. Results of the primary validation only are discussed here, with special emphasis on a comparison of findings for the white and Negro subgroups in the total sample. These results are shown in Table 2.

In Table 2, the left-hand column of figures (R) under each criterion measure gives the raw estimates obtained for the total sample and for the white and Negro subgroups. However, these estimates are based on different sample sizes, as shown by the number of cases listed in column "N," and on different numbers of test battery variables contributing to the correlation, as shown in column "m" for each criterion measure. Such raw estimates tend to be inflated, especially when using small samples utilizing a disproportionately large number of variables. A correction was made for this bias (Guilford, 1956, p. 399) and yielded the corrected estimates given in the second column (R').

In general, slightly higher coefficients were obtained for the Wave II sample. However, these increases are not statistically significant, and the results from Wave I were essentially replicated in Wave II. The tentative conclusions given below apply to both samples.

The final unbiased estimates of the multiple correlations are at generally acceptable levels, after allowing for the expected decreases when they are based

CHAPTER VII -- Table 1

Test (Predictor) Variables for Wave I and Wave II Samples of Patrolmen

AREA OF MEASUREMENT	TEST USED	NUMBER OF VARIABLES	WAVE I SAMPLE	WAVE II SAMPLE	COMMON TESTS
MOTIVATION					
Background & Experience	PHI	15	XX	XX	XX
Work Interest	WII	14		XX	
INTELLECTUAL SKILL					
Reasoning	NVR	1	XX		
Language	UC	1	XX		
Perception	CS	1	XX	XX	XX
	CF	1	XX	XX	XX
	PS	1	XX	XX	XX
APTITUDE					
Creative	AC	2	XX		
	CREE	14		XX	
Social	SI	8	XX	XX	XX
BEHAVIOR					
Temperament	TC	24	XX	XX	XX
	EPPS	15		XX	
Pressure Tolerance	PRESS	5	XX	XX	XX
Personality Systems	BI	1	XX		
	HAND	8	XX		
	MPI	3	XX		
	A-D	7	XX	XX	XX
TOTAL NUMBER OF TESTS			14	11	8
ADMINISTRATION TIME (HOURS)			4	4	3

CHAPTER VII -- Table 2

Multiple Correlations of Test Batteries with Criterion Measures

CRITERION MEASURES

TEST BATTERIES	Groups	N	PC Rating			1966 CPD Rating			CPD Tenure			CPD Awards			IID Complaints			Disciplinary Actions			Number of Arrests 1966			Times Absent		
			R	R'	m	R	R'	m	R	R'	m	R	R'	m	R	R'	m	R	R'	m	R	R'	m	R	R'	m
WAVE I SAMPLE																										
Wave I Tests	Total	175	.60	.50	25	.67	.61	23	.72	.67	21	.57	.48	20	.53	.46	14	.47	.38	16	.53	.46	15	.55	.43	25
	White	128	.62	.53	19	.66	.57	22	.74	.67	22	.59	.47	20	.64	.53	24	.51	.38	18	.63	.51	24	.67	.55	26
	Negro	47	.92	.85	22	.90	.83	17	.92	.88	13	.85	.74	18	.91	.82	21	.89	.75	24	.84	.71	19	.96	.83	33
Tests in Common to Both Samples	Total	175	.58	.49	23	.65	.59	20	.67	.63	16	.47	.39	14	.48	.41	14	.44	.36	13	.45	.38	12	.50	.40	18
	White	128	.61	.51	20	.63	.56	16	.68	.62	16	.47	.37	12	.58	.46	19	.43	.33	11	.54	.47	12	.59	.49	19
	Negro	47	.90	.80	21	.94	.87	24	.99	.96	33	.86	.75	18	.92	.80	27	.93	.85	23	.94	.86	24	.85	.73	18
WAVE II SAMPLE																										
Wave II Tests	Total	192	.67	.60	27	.63	.55	27	.81	.78	19	.57	.50	20	.61	.52	28	.66	.57	33	.54	.47	19	.60	.50	26
	White	126	.77	.65	37	.69	.56	29	.84	.81	16	.60	.53	15	.65	.53	24	.77	.66	33	.55	.44	17	.74	.63	32
	Negro	66	.98	.93	41	.88	.79	26	.97	.93	37	.88	.79	28	.95	.89	36	.88	.77	29	.88	.80	24	.95	.86	43
Tests in Common to Both Samples	Total	192	.54	.48	16	.50	.43	17	.80	.78	15	.49	.41	17	.51	.43	16	.58	.51	19	.44	.37	12	.38	.29	13
	White	126	.67	.57	24	.59	.48	19	.84	.81	14	.45	.36	11	.57	.50	13	.65	.57	19	.44	.33	12	.50	.41	13
	Negro	66	.81	.73	18	.93	.85	35	.92	.88	20	.87	.79	25	.85	.76	23	.86	.75	26	.74	.62	17	.75	.60	21

on the restricted test battery common to both samples. For the paired-comparison performance rating, which is the major criterion measure in the study, the unbiased correlation estimates are all statistically significant at the .01 level of confidence or better. Throughout, however, the highest correlations were obtained for the tenure criterion. Thus, for the two samples of beat patrolmen, the test battery scores seem to be significantly related both to level of job performance and to tenure.

When the multiple correlations are calculated separately for the two racial subgroups, that obtained for the white subgroup increases slightly over the estimate obtained for the total sample, and that for the Negro subgroup more spectacularly. A t-test of the significance of the difference between the corrected correlation coefficients obtained for the two subgroups indicates that for both the Wave I and the Wave II samples the difference is significant beyond the .01 level of confidence. These results underline the necessity for a general procedure of separate validations for different racial groups. With regard to this particular study, the results indicate that the test battery has higher validity for the Negro subgroup. An explanation of the different validity estimates obtained for the white and Negro subgroups was sought in a comparison of their scores on both the test and the criterion measures.

RESPONSE OF WHITE AND NEGRO PATROLMEN TO THE TEST (PREDICTOR) VARIABLES

Responses of the white and Negro subgroups will be discussed for each of the major areas of the test battery. The first of these is personal background and experience.

Personal Background and Experience

Objective measures of dimensions of personal background are obtained through the Personal History Index. Table 3 shows the mean scores for the subgroups on each of the 15 dimensions of the instrument, the standard deviations of these means (S.D.), and the probability (p) of the significance of the difference

CHAPTER VII -- Table 3

T-Tests for the Significance of the Differences between the Mean Scores for White and Negro Subgroups on Dimensions of the Personal History Index

Wave I Sample

Variable	Dimension	White Group (N = 170)		Negro Group (N = 62)			
		Mean	S.D.	Mean	S.D.	t	p
1	School Achievement	1.55	1.24	2.16	1.49	2.85	.005**(N)
2	Higher Educational Achievement	3.28	1.64	3.00	1.39	1.29	.194
3	Drive	4.14	1.83	3.90	1.73	0.88	.619
4	Leadership & Group Participation	2.14	1.57	2.34	1.57	0.84	.596
5	Financial Responsibility	2.36	1.40	2.16	1.25	1.02	.307
6	Early Family Responsibility	6.19	2.81	6.10	2.55	0.25	.799
7	Parental Family Adjustment	6.34	2.37	5.73	2.07	1.89	.056
8	Stability	4.01	2.40	4.27	2.18	0.80	.571
9	School Activities	2.36	1.60	2.76	1.36	1.87	.060
10	Professional Successful Parents	3.33	1.45	3.00	1.26	1.68	.090
11	Educational-Vocational Consistency	1.79	0.86	1.73	0.90	0.47	.644
12	Vocational Decisiveness	2.05	1.11	2.06	1.23	0.10	.919
13	Vocational Satisfaction	4.88	1.78	4.13	1.90	2.67	.008**(W)
14	Selling Experience	5.28	1.54	5.03	1.59	1.04	.300
15	General Health	2.59	1.18	2.63	1.10	0.24	.803

Wave II Sample

Variable	Dimension	White Group (N = 188)		Negro Group (N = 92)			
		Mean	S.D.	Mean	S.D.	t	p
1	School Achievement	1.31	1.14	2.41	1.50	6.21	.000***(N)
2	Higher Educational Achievement	2.86	1.48	2.72	1.56	0.73	.530
3	Drive	4.39	1.84	3.83	1.79	2.44	.015*(W)

CHAPTER VII -- Table 3 cont.

4	Leadership & Group Participation	2.05	1.44	2.15	1.43	0.57	.577
5	Financial Responsibility	2.10	1.36	1.65	1.15	2.88	.005**(W)
6	Early Family Responsibility	5.49	3.02	4.74	2.51	2.20	.027*(W)
7	Parental Family Adjustment	6.30	2.23	5.49	2.28	2.79	.006**(W)
8	Stability	3.31	2.37	3.15	2.03	0.57	.577
9	School Activities	2.34	1.60	2.46	1.59	0.60	.559
10	Professional Successful Parents	3.29	1.46	3.13	1.44	0.88	.615
11	Educational-Vocational Consistency	1.71	0.83	1.66	0.85	0.46	.651
12	Vocational Decisiveness	2.05	1.20	2.07	1.20	0.08	.935
13	Vocational Satisfaction	4.93	1.77	4.75	2.03	0.70	.511
14	Selling Experience	5.29	1.59	4.68	1.66	2.88	.005**(W)
15	General Health	2.67	1.20	2.83	1.07	1.10	.274

Total Sample

Variable	Dimension	White Group (N = 358)		Negro Group (N = 154)			
		Mean	S.D.	Mean	S.D.	t	p
1	School Achievement	1.42	1.20	2.31	1.50	6.48	.000***(N)
2	Higher Educational Achievement	3.06	1.57	2.83	1.50	1.56	.114
3	Drive	4.27	1.84	3.86	1.76	2.38	.017*(W)
4	Leadership & Group Participation	2.09	1.50	2.23	1.49	0.94	.648
5	Financial Responsibility	2.22	1.39	1.86	1.21	2.99	.003**(W)
6	Early Family Responsibility	5.83	2.94	5.29	2.61	2.06	.037*(W)
7	Parental Family Adjustment	6.32	2.29	5.58	2.20	3.39	.001**(W)
8	Stability	3.64	2.41	3.60	2.16	0.17	.863
9	School Activities	2.35	1.60	2.58	1.51	1.56	.116
10	Professional Successful Parents	3.31	1.46	3.08	1.37	1.72	.082
11	Educational-Vocational Consistency	1.75	0.84	1.69	0.87	0.72	.523
12	Vocational Decisiveness	2.05	1.15	2.06	1.21	0.13	.894
13	Vocational Satisfaction	4.90	1.77	4.50	2.00	2.15	.030*(W)
14	Selling Experience	5.28	1.56	4.82	1.64	2.93	.004**(W)
15	General Health	2.63	1.19	2.75	1.08	1.07	.285

* $p < .05$
** $p < .01$
*** $p < .001$
(N) = Negro group mean higher
(W) = White group mean higher

between the means of the two groups. One, two, and three asterisks indicate significance levels of .05, .01, and .001 respectively. A superscript W or N following the final asterisk indicates whether the white or the Negro group mean is higher.

The first eight dimensions of the Index are generally the most predictive and the ones scored for operational use. For Wave I, the only significant difference on these eight dimensions occurs in favor of the Negro group on School Achievement. This result is repeated for Wave II, with the addition of significant differences in favor of the white group on dimensions dealing with personal drive, financial responsibility, and family responsibility and adjustment. Results for the total sample--the sum or combination of Wave I and Wave II--are the same as those obtained for Wave II.

It seems a tenable hypothesis that the Negro group's generally lower scores in this background area are the result of educational, vocational, and social deprivation. This hypothesis would also explain the single instance in which the Negro group scored significantly higher--the dimension of School Achievement. If educational standards are generally lower in predominantly Negro schools, the Negro patrolmen applicants would need a higher educational grade level to pass the civil service screening procedure than would their white counterparts. It is possible that in this study we are dealing with a relatively more highly selected Negro group.

Work Interests

The Work Interest Index (a non-verbal, check-mark-response instrument) measures the strength of interests in 12 work areas, as well as the flexibility of these interests and the individual's vocational aspiration level. Table 4 shows these results. Except for the white group's stronger interest in occupations designated "Authority & Prestige," all significant differences were in favor of the Negro group, which showed higher interest in social, creative and artistic, clerical, and theoretical and scientific occupations.

CHAPTER VII -- Table 4

T-Tests for the Significance of the Differences between the Mean Scores for White and Negro Subgroups on Dimensions of the Work Interest Index

		Wave II Sample					
		White Group (N = 188)		Negro Group (N = 92)			
Variable	Dimension	Mean	S.D.	Mean	S.D.	t	p
641	Professional & Technical	3.86	2.69	4.03	2.70	0.51	.616
642	Social & Verbal	4.94	2.19	5.57	2.14	2.26	.023*(N)
643	Authority & Prestige	5.89	1.59	5.47	1.66	2.01	.042*(W)
644	Artistic & Interpretative	2.49	2.34	3.13	2.34	2.13	.032*(N)
645	Artistic & Stylized	2.23	2.28	1.92	2.10	1.10	.270
646	Artistic & Creative	3.34	2.67	4.37	2.88	2.88	.005**(N)
647	Technical & Scientific	4.34	2.24	5.12	2.24	2.75	.006**(N)
648	Clerical & Routine	2.57	2.59	3.80	2.96	3.40	.001**(N)
649	Business Contact & Structured	1.76	2.36	2.34	2.56	1.82	.066
650	Personal Service & Persuasive	3.26	2.10	3.15	2.06	0.39	.700
651	Mechanical & Productive	4.39	2.89	4.15	2.93	0.63	.535
652	Control of Massive Equipment	4.09	2.98	3.68	2.93	1.08	.282
653	FLEXIBILITY	43.14	18.46	46.74	18.52	1.52	.125
654	ASPIRATION LEVEL	28.29	2.29	28.12	2.11	0.62	.543

* $p < .05$
** $p < .01$

(N) = Negro group mean higher
(W) = White group mean higher

The Work Interest Index was administered only to the Wave II sample. Further discussion of the different vocational interests of the two groups will be presented later.

Intellectual Skill

Five tests, each designed to measure a specific mental ability, were administered to the Wave I sample. As shown in Table 5, the white group scored higher on all five tests and significantly higher on three of them. These results accord with those obtained in other studies of this nature. Only one of the tests--Understanding Communication, which deals specifically with language facility--is verbal. The remaining four are non-verbal, and on these four, there appears to be a fairly systematic relationship between significance level and the time limit for the test. Arranging them from longest to shortest time limit, we find the following:

Test	Time Limit	Probability Value
Non-Verbal Reasoning	Untimed, approx. 20 mins.	.09
Closure Flexibility	15 minutes	.10
Perceptual Speed	5 minutes	.02
Closure Speed	3 minutes	.00

With the three tests which were retained in the battery--Closure Flexibility, Perceptual Speed, and Closure Speed--this relationship repeated for the Wave II sample and for the total sample. This result suggests that the more speeded the test, the greater the likelihood that the Negro group will have a significantly lower mean score. In view of this tentative hypothesis, the Press Test, ordinarily included in the "Behavior" section of the battery, will be discussed next.

Pressure Tolerance

The three parts of the Press Test measure reaction time to verbal stimuli (Part I), to color stimuli (Part II), and to color stimuli under pressure of verbal distraction (Part III). There are also two derived scores, which contrast the individual's own reaction times to verbal and to color stimuli (Part I minus

CHAPTER VII -- Table 5

T-Tests for the Significance of the Differences between the Mean Scores for White and Negro Subgroups on Tests of Specific Mental Abilities

Wave I Sample

Variable	Test	White Group N	White Group Mean	White Group S.D.	Negro Group N	Negro Group Mean	Negro Group S.D.	t	p
16	Non-Verbal Reasoning	170	31.89	5.23	62	30.63	4.83	1.71	.085
17	Understanding Communication	170	29.75	5.01	62	28.00	5.66	2.14	.032*(W)
18	Closure Speed	170	12.65	5.13	62	10.26	4.13	3.63	.001***(W)
19	Closure Flexibility	170	54.15	26.72	62	48.37	22.25	1.64	.098
20	Perceptual Speed	160	51.68	13.83	61	47.11	12.27	2.37	.018*(W)

Wave II Sample

Variable	Test	White Group N	White Group Mean	White Group S.D.	Negro Group N	Negro Group Mean	Negro Group S.D.	t	p
18	Closure Speed	188	12.92	4.69	92	11.14	4.56	3.02	.003**(W)
19	Closure Flexibility	185	59.05	24.28	91	56.88	24.53	0.69	.502
20	Perceptual Speed	186	53.32	12.13	90	50.06	15.32	1.76	.076

Total Sample

Variable	Test	White Group N	White Group Mean	White Group S.D.	Negro Group N	Negro Group Mean	Negro Group S.D.	t	p
18	Closure Speed	358	12.79	4.91	154	10.79	4.42	4.55	.000***(W)
19	Closure Flexibility	355	56.70	25.59	153	53.43	24.00	1.38	.166
20	Perceptual Speed	346	52.56	12.97	151	48.87	14.24	2.72	.007**(W)

* $p < .05$

** $p < .01$

*** $p < .001$

(W) = White group mean higher

Part II) and the individual's own reaction times to color stimuli with and without stress (Part II minus Part III). The Press Test is the most speeded one in the battery, the time limit for each of the three parts being exactly 90 seconds.

As shown in Table 6, the test yields consistent and significant differences between the racial group means for all three samples. The white group scores significantly higher on each of the three separate parts and on reaction time under pressure (Part II minus Part III). In contrast, the Negro group has a significantly higher mean score for Part I minus Part II. In other words, the Negro group shows a relatively faster reaction to color than to verbal stimuli than does the white group. The direct scores for reaction time to verbal stimuli, and to color stimuli with and without stress, support the suggestion made in the previous section that Negro group scores compare most poorly with those of the white group in speeded or pressure test situations.

Aptitude

This section of the test battery covers two broad areas of behavior--the creative potential of the individual and insight in social situations. Table 7 shows results for the two tests of creative potential, Table 8 results for the Test of Social Insight. In contrast to findings on the background and ability measures, a far smaller proportion of the dimensions yield significant differences between the racial group means. Furthermore, such significant differences as do exist are in favor of the Negro group.

On the total scores for these tests, there are no significant differences between the racial groups. However, certain dimensions of the Cree Questionnaire show the Negro group to have significantly greater interest in the artistic, scientific, and theoretical occupations. These results accord with those obtained from the Work Interest Index. As further points, the Negro group is less involved in the work situation and has a greater tendency to withdraw in social situations.

CHAPTER VII -- Table 6

T-Tests for the Significance of the Differences
between the Mean Scores for White and Negro Subgroups
on a Test of Pressure Tolerance

Wave I Sample

		White Group			Negro Group				
Variable	PRESS TEST	N	Mean	S.D.	N	Mean	S.D.	t	p
25	Part I	165	73.52	16.56	62	70.34	16.63	1.28	.200
26	Part II	165	70.79	15.41	62	62.37	14.93	3.73	.000***(W)
27	Part III	164	57.44	15.34	62	52.74	13.03	2.28	.022*(W)
28	Part I - II	165	3.47	8.90	62	8.29	13.37	2.61	.010**(N)
29	Part II - III	164	13.38	10.26	62	9.63	8.76	2.72	.007**(W)

Wave II Sample

		White Group			Negro Group				
Variable	PRESS TEST	N	Mean	S.D.	N	Mean	S.D.	t	p
25	Part I	188	77.71	14.92	92	70.68	17.31	3.32	.001**(W)
26	Part II	188	74.89	14.19	92	66.78	15.72	4.16	.000***(W)
27	Part III	188	60.79	13.49	91	55.22	14.10	3.12	.002**(W)
28	Part I - II	188	3.12	9.10	92	4.79	7.58	1.61	.104
29	Part II - III	188	14.02	9.60	91	11.22	8.98	2.38	.017*(W)

Total Sample

		White Group			Negro Group				
Variable	PRESS TEST	N	Mean	S.D.	N	Mean	S.D.	t	p
25	Part I	353	75.75	15.85	154	70.55	17.04	3.22	.002**(W)
26	Part II	353	72.97	14.92	154	65.01	15.56	5.35	.000***(W)
27	Part III	352	59.23	14.48	153	54.22	13.73	3.70	.000***(W)
28	Part I - II	353	3.28	9.01	154	6.20	10.45	3.00	.003**(N)
29	Part II - III	352	13.72	9.92	153	10.58	8.93	3.51	.001***(W)

* $p < .05$
** $p < .01$
*** $p < .001$
(W) = White group mean higher
(N) = Negro group mean higher

CHAPTER VII -- Table 7

T-Tests for the Significance of the Differences between the Mean Scores for White and Negro Subgroups on Tests of Creative Potential

AC Test of Creative Ability

Wave I Sample

Variable	Dimension	White Group N	White Group Mean	White Group S.D.	Negro Group N	Negro Group Mean	Negro Group S.D.	t	p
23	Part II	169	11.44	5.23	62	10.66	4.48	1.11	.269
24	Part III	170	24.44	9.43	62	23.79	8.39	0.50	.622

Cree Questionnaire

Wave II Sample

Variable	Dimension	White Group (N = 188) Mean	White Group (N = 188) S.D.	Negro Group (N = 92) Mean	Negro Group (N = 92) S.D.	t	p
656	Total Score	66.56	12.79	67.68	12.24	0.71	.512
657	Dominance	5.13	3.23	5.05	2.82	0.19	.841
658	Indifference	3.63	1.73	3.92	1.75	1.30	.190
659	Independence	5.55	2.13	6.27	2.28	2.54	.011*(N)
660	Unstructured Situation	3.54	2.04	3.38	1.89	0.63	.534
661	Unsystematic Selective Activity	4.59	1.90	4.97	2.14	1.43	.150
662	Personally Involved Attitude	3.72	1.98	3.62	2.00	0.39	.702
663	Pressure Situation	2.20	1.35	2.08	1.17	0.77	.549
664	High Energy Level	5.34	2.30	4.83	2.30	1.75	.077
665	Fast Reaction	5.34	1.75	4.96	1.78	1.70	.087
666	High Ideational Spontaneity	4.11	1.64	4.50	1.59	1.89	.057
667	High Scientific & Theoretical	3.59	2.14	4.14	2.13	2.04	.040*(N)
668	High Artistic	2.22	1.61	2.77	1.73	2.56	.011*(N)
669	High Mechanical	3.56	1.76	3.17	1.73	1.73	.081

* $p < .05$

(N) = Negro group mean higher

CHAPTER VII -- Table 8

T-Tests for the Significance of the Differences between the Mean Scores for White and Negro Subgroups on the Test of Social Insight

Wave I Sample

Variable	Dimension	White Group (N = 170)		Negro Group (N = 62)		t	p
		Mean	S.D.	Mean	S.D.		
88	Withdrawal	4.58	2.36	5.03	2.28	1.31	.189
89	Passivity	8.71	3.31	8.92	3.36	0.43	.673
90	Cooperativeness	33.62	4.63	32.94	4.17	1.06	.289
92	Competitiveness	6.32	2.24	6.56	2.17	0.75	.542
94	Aggression	6.76	2.64	6.60	2.83	0.39	.699
95	Total	22.81	5.12	23.05	5.32	0.30	.762

Wave II Sample

Variable	Dimension	White Group (N = 185)		Negro Group (N = 90)		t	p
		Mean	S.D.	Mean	S.D.		
88	Withdrawal	4.31	2.02	5.00	2.20	2.48	.013*(N)
89	Passivity	9.09	3.41	8.60	3.61	1.07	.284
90	Cooperativeness	33.31	4.87	33.37	4.64	0.10	.920
92	Competitiveness	5.94	2.15	6.01	2.12	0.28	.779
94	Aggression	7.31	2.64	6.97	2.59	1.01	.313
95	Total	22.56	4.94	22.36	4.83	0.32	.748

CHAPTER VII -- Table 8 cont.

Total Sample

Variable	Dimension	White Group (N = 355) Mean	White Group (N = 355) S.D.	Negro Group (N = 152) Mean	Negro Group (N = 152) S.D.	t	p
88	Withdrawal	4.44	2.19	5.01	2.23	2.64	.008**(N)
89	Passivity	8.91	3.37	8.73	3.52	0.52	.607
90	Cooperativeness	33.46	4.76	33.19	4.46	0.60	.556
92	Competitiveness	6.12	2.20	6.24	2.16	0.56	.582
94	Aggression	7.05	2.66	6.82	2.70	0.88	.616
95	Total	22.68	5.03	22.64	5.05	0.08	.003

* $p < .05$

** $p < .01$

(N) = Negro group mean higher

Behavior

The Temperament Comparator (Tables 9 and 10) was administered to both patrolman samples. This instrument was developed by the Industrial Relations Center to measure some of the relatively permanent behavior attributes of the individual. It yields comparative measures of 18 temperament traits and five temperament factors. The latter were determined by factor analysis, and each factor score is a composite of five trait scores. These factor scores are therefore the most reliable ones derived from this instrument. Racial group means for the factor scores differed significantly only for the factor Controlled. The Negro group mean was lower, indicating greater impulsiveness, demonstrativeness, and behavioral variability. The higher white group mean indicates greater caution, control, and perseverance. Group differences on the trait scores were significant for some of the traits contributing to the factor Controlled. Thus, the Negro group scores higher on Demonstrative, the white group on Persevering and Steady Worker. In addition, the Negro group scores higher on Talkative and Emotionally Stable, the white group on Energetic, Enthusiastic, and Quick Worker.

On four other personality questionnaires, the two groups are relatively similar, as shown in Table 11. In fact, the group means differ significantly on only two out of a total of 28 widely divergent behavior dimensions. One of these is a composite score from selected MMPI items. The other is the Intraception score on the Edwards Personal Preference Schedule. The higher score of the Negro group on the latter indicates that they are better able to analyze the feelings and behavior of others and to put themselves in the place of others.

The final personality measure, the Arrow-Dot Test (Table 12), differs from the other, untimed instruments in this area in that it is a high-speed, objective, performance test. A full description of it is given in Chapter IV. It provides three scores. One is called the Id score, defined as a measure of the tendency to uncontrolled, impulsive behavior. The second is the Ego score, which essentially is a measure of the ability to function realistically and of the degree of

CHAPTER VII -- Table 9

T-Tests for the Significance of the Differences
between the Mean Scores for White and Negro Subgroups
on the Factor Scores from the Temperament Comparator

Wave I Sample

Variable	Factor	White Group (N = 148) Mean	White Group (N = 148) S.D.	Negro Group (N = 54) Mean	Negro Group (N = 54) S.D.	t	p
31	Controlled	54.00	10.27	50.07	8.98	2.62	.009**(W)
32	Stable	54.48	11.65	54.26	12.71	0.11	.908
33	Self-Reliant	52.95	11.92	54.50	14.30	0.71	.513
34	Excitable	37.48	12.21	35.39	11.37	1.12	.261
35	Sociable	43.39	11.04	43.74	9.42	0.22	.820

Wave II Sample

Variable	Factor	White Group (N = 182) Mean	White Group (N = 182) S.D.	Negro Group (N = 85) Mean	Negro Group (N = 85) S.D.	t	p
31	Controlled	52.26	9.28	50.51	10.80	1.28	.198
32	Stable	54.96	11.41	55.36	11.70	0.26	.789
33	Self-Reliant	55.25	12.32	56.07	11.89	0.52	.611
34	Excitable	37.13	12.28	34.79	13.03	1.39	.163
35	Sociable	43.51	12.89	45.98	9.24	1.78	.073

Total Sample

Variable	Factor	White Group (N = 330) Mean	White Group (N = 330) S.D.	Negro Group (N = 139) Mean	Negro Group (N = 139) S.D.	t	p
30	Consistency	53.99	36.92	50.42	40.91	0.89	.621
31	Controlled	53.04	9.78	50.34	10.13	2.66	.008**(W)
32	Stable	54.75	11.52	54.94	12.11	0.16	.870
33	Self-Reliant	54.22	12.20	55.46	12.90	0.97	.664
34	Excitable	37.29	12.25	35.02	12.42	1.81	.068
35	Sociable	43.45	12.10	45.11	9.38	1.59	.108

** $p < .01$
(W) = White group mean higher

CHAPTER VII -- Table 10

T-Tests for the Significance of the Differences between the Mean Scores for White and Negro Subgroups on the Trait Scores from the Temperament Comparator

		Wave I Sample					
		White Group (N = 148)		Negro Group (N = 54)			
Variable	Temperament Trait	Mean	S.D.	Mean	S.D.	t	p
36	Calm	10.54	4.05	10.07	3.98	0.73	.526
37	Cautious	10.80	4.08	10.67	4.47	0.20	.839
38	Decisive	8.81	3.72	8.69	4.03	0.20	.837
39	Demonstrative	3.77	3.49	5.20	3.70	2.45	.014*(N)
40	Emotionally Stable	9.86	3.90	10.72	3.67	1.43	.149
41	Energetic	8.44	3.59	8.22	3.67	0.37	.712
42	Enthusiastic	8.59	3.53	7.96	3.39	1.15	.250
43	Even-Tempered	9.97	4.62	11.00	4.23	1.49	.135
44	Lively	6.29	3.53	6.93	3.91	1.04	.300
45	Persevering	6.99	4.23	5.61	3.77	2.20	.027*(W)
46	Prompt Starter	8.43	4.28	8.50	4.34	0.10	.919
47	Quick Worker	7.22	4.05	6.46	4.06	1.16	.246
48	Seeks Company	5.25	4.32	4.06	4.38	1.71	.085
49	Self-Confident	10.79	4.26	11.81	4.22	1.51	.128
50	Serious	10.94	3.84	10.59	4.52	0.50	.625
51	Socially At Ease	9.01	3.86	8.63	3.67	0.63	.535
52	Steady Worker	11.97	3.06	11.22	3.62	1.35	.177
53	Talkative	5.26	4.58	6.67	5.77	1.60	.108
30	Consistency	55.46	40.91	46.04	36.81	1.55	.119

CHAPTER VII -- Table 10 cont.
Wave II Sample

Variable	Temperament Trait	White Group (N = 182) Mean	White Group (N = 182) S.D.	Negro Group (N = 85) Mean	Negro Group (N = 85) S.D.	t	p
36	Calm	10.19	4.04	10.86	3.82	1.31	.188
37	Cautious	10.59	4.04	10.82	4.09	0.44	.666
38	Decisive	9.12	3.95	9.79	3.72	1.33	.181
39	Demonstrative	4.01	3.60	4.93	3.85	1.84	.063
40	Emotionally Stable	10.45	3.80	11.08	3.67	1.30	.191
41	Energetic	8.77	3.83	7.62	3.80	2.28	.022*(W)
42	Enthusiastic	8.85	3.50	7.87	3.64	2.06	.038*(W)
43	Even-Tempered	10.76	4.33	10.53	4.12	0.41	.683
44	Lively	6.72	3.81	6.73	3.69	0.02	.982
45	Persevering	6.79	3.81	6.11	3.81	1.35	.175
46	Prompt Starter	7.73	4.36	7.05	3.86	1.29	.196
47	Quick Worker	7.32	3.98	6.26	3.41	2.23	.025*(W)
48	Seeks Company	4.54	4.42	4.49	3.92	0.08	.932
49	Self-Confident	11.47	4.19	12.00	3.69	1.04	.301
50	Serious	10.90	3.81	11.25	4.23	0.64	.531
51	Socially At Ease	8.54	4.01	9.08	4.12	1.00	.680
52	Steady Worker	11.01	3.60	10.11	3.91	1.79	.071
53	Talkative	5.27	4.42	6.42	5.48	1.69	.088
30	Consistency	52.80	33.27	53.20	43.09	0.07	.938

CHAPTER VII -- Table 10 cont.

		Total Sample					
		White Group (N = 330)		Negro Group (N = 139)			
Variable	Temperament Trait	Mean	S.D.	Mean	S.D.	t	p
36	Calm	10.35	4.05	10.55	3.90	0.52	.609
37	Cautious	10.68	4.06	10.76	4.24	0.18	.849
38	Decisive	8.98	3.85	9.36	3.88	0.96	.662
39	Demonstrative	3.90	3.56	5.04	3.80	3.00	.003**(N)
40	Emotionally Stable	10.18	3.86	10.94	3.67	2.00	.043*(N)
41	Energetic	8.62	3.73	7.86	3.76	2.01	.042*(W)
42	Enthusiastic	8.73	3.51	7.91	3.55	2.40	.020*(W)
43	Even-Tempered	10.40	4.48	10.71	4.17	0.71	.518
44	Lively	6.53	3.69	6.81	3.78	0.73	.528
45	Persevering	6.88	4.01	5.91	3.80	2.45	.014*(W)
46	Prompt Starter	8.05	4.34	7.61	4.11	1.02	.307
47	Quick Worker	7.27	4.01	6.34	3.67	2.44	.014*(W)
48	Seeks Company	4.86	4.39	4.32	4.11	1.26	.207
49	Self-Confident	11.17	4.24	11.93	3.91	1.87	.058
50	Serious	10.92	3.82	10.99	4.36	0.18	.856
51	Socially At Ease	8.75	3.95	8.91	3.96	0.39	.702
52	Steady Worker	11.44	3.40	10.54	3.84	2.39	.016*(W)
53	Talkative	5.27	4.49	6.52	5.60	2.33	.019*(N)
30	Consistency	53.99	36.92	50.42	40.91	0.89	.621

* $p < .05$
** $p < .01$

(N) = Negro group mean higher
(W) = White group mean higher

CHAPTER VII -- Table 11

T-Tests for the Significance of the Differences between the Mean Scores for White and Negro Subgroups on Four Personality Inventories

Edwards Personal Preference Schedule

Wave II Sample

Variable	Dimension	White Group (N = 188) Mean	White Group (N = 188) S.D.	Negro Group (N = 92) Mean	Negro Group (N = 92) S.D.	t	p
670	Achievement	16.21	4.02	16.82	3.92	1.19	.232
671	Deference	13.26	3.64	13.53	3.69	0.59	.563
672	Order	12.90	4.40	13.95	4.64	1.79	.070
673	Exhibition	14.69	3.76	14.07	3.97	1.24	.212
674	Autonomy	13.10	4.22	13.96	4.43	1.53	.122
675	Affiliation	12.65	3.89	11.65	4.38	1.84	.063
676	Intraception	15.34	4.57	17.28	4.08	3.58	.001***(N)
677	Succorance	9.40	4.11	8.89	4.24	0.95	.654
678	Dominance	16.24	4.65	16.39	3.81	0.29	.770
679	Abasement	12.12	5.11	11.97	7.98	0.17	.860
680	Nurturance	13.30	4.68	12.58	4.53	1.23	.218
681	Change	15.28	4.57	15.05	4.42	0.40	.693
682	Endurance	16.53	4.73	16.41	5.03	0.19	.845
683	Heterosexuality	15.81	6.42	16.59	6.76	0.92	.636
684	Aggression	13.18	4.03	12.85	5.14	0.53	.600
685	Consistency	4.23	5.03	3.82	2.34	0.94	.647

Behavior Inventory

Wave I Sample

Variable	Dimension	White Group (N = 170) Mean	White Group (N = 170) S.D.	Negro Group (N = 62) Mean	Negro Group (N = 62) S.D.	t	p
87	Total Score	13.62	2.82	15.06	2.66	3.57	.001***(N)

CHAPTER VII -- Table 11 cont.

Hand Test

Wave I Sample

Variable	Dimension	White Group (N = 170) Mean	White Group (N = 170) S.D.	Negro Group (N = 62) Mean	Negro Group (N = 62) S.D.	t	p
70	Interpersonal Response	9.12	2.94	9.27	3.23	0.32	.749
71	Environmental Response	6.74	2.61	6.69	3.02	0.11	.909
72	Maladjusted Response	2.38	1.78	2.53	1.57	0.62	.546
73	Withdrawal Response	0.65	1.09	1.16	2.48	1.55	.118
78	Inner Directed	4.90	2.31	5.32	2.47	1.17	.243
79	Outer Directed	3.48	1.84	3.11	1.75	1.39	.161
80	Acting Out Ratio	1.75	1.43	1.92	1.47	0.78	.560
83	Pathological Score	3.69	2.85	4.90	5.07	1.77	.048*(N)

Maudsley Personality Inventory

Wave I Sample

Variable	Dimension	White Group (N = 170) Mean	White Group (N = 170) S.D.	Negro Group (N = 62) Mean	Negro Group (N = 62) S.D.	t	p
84	Extroversion	29.62	7.78	29.21	6.62	0.40	.693
85	Neuroticism	18.04	10.49	16.97	8.86	0.77	.552
86	Undecided (?)	1.99	2.50	2.03	2.91	0.10	.913

* $p < .05$
*** $p < .001$

(N) = Negro group mean higher

CHAPTER VII -- Table 12

T-Tests for the Significance of the Differences
between the Mean Scores for White and Negro Subgroups
on the Arrow-Dot Test

Wave I Sample

Variable	Dimension	White Group (N = 170) Mean	White Group (N = 170) S.D.	Negro Group (N = 62) Mean	Negro Group (N = 62) S.D.	t	p
96	Id	4.51	4.66	5.02	4.79	0.72	.519
97	Ego	20.48	6.47	19.47	6.11	1.09	.277
98	Superego	4.24	3.30	4.89	3.50	1.25	.208

Wave II Sample

Variable	Dimension	White Group (N = 188) Mean	White Group (N = 188) S.D.	Negro Group (N = 92) Mean	Negro Group (N = 92) S.D.	t	p
96	Id	3.80	4.34	6.27	6.65	3.22	.002**(N)
97	Ego	21.79	6.14	18.78	7.84	3.22	.002**(W)
98	Superego	4.04	3.45	4.61	3.60	1.26	.207

Total Sample

Variable	Dimension	White Group (N = 358) Mean	White Group (N = 358) S.D.	Negro Group (N = 154) Mean	Negro Group (N = 154) S.D.	t	p
96	Id	4.14	4.51	5.77	6.00	3.01	.003**(N)
97	Ego	21.16	6.33	19.05	7.20	3.15	.002**(W)
98	Superego	4.13	3.38	4.72	3.56	1.73	.080

** $p < .01$
(N) = Negro group mean higher
(W) = White group mean higher

ego strength and control. Finally, there is the Superego score, a measure of superego strength, especially in circumstances where its demands are not consciously experienced but are viewed as external restrictions. For Wave II and for the total sample, there are significant differences between the group means. The Negro group scores higher on the Id and Superego measures, the white group higher on the Ego score. These differential behavior characteristics accord with other findings of the battery. However, they must be regarded with caution in view of the evidence that differences between the two racial groups may be exaggerated in high-speed or pressure test situations.

Summary and Discussion of the Response of White and Negro Patrolmen to the Test (Predictor) Variables

This analysis has concentrated on obtained differences of response, and thus tends to overstress such differences. Actually, the two racial groups are similar on more dimensions than they are different. Another important fact to be borne in mind is that the distributions for the two groups overlap on all variables. That is, if a Negro group mean is the lower, there will be some Negroes who score higher than most of the whites. Conversely, if the Negro group mean is the higher, there will be some whites who score higher than most of the Negroes.

With these considerations in mind, we can say that in the background area, the Negroes show higher educational achievement but are lower in the dimensions of personal drive, financial responsibility, and family responsibility and adjustment.

In the work interest area, the Negro group are consistently high in their interest in creative and artistic-interpretative occupations. There is also some evidence of higher interest in the theoretical, scientific, and clerical areas. On the other hand, the white group showed higher interest only in the prestige and status occupations.

There were no generalized differences in either creative potential or social insight, but the Negro group scores lower on all mental ability tests and most markedly so on the speeded and pressure tests.

In the behavior area, the Negro group is more impulsive, demonstrative, and generally outgoing. There is a more immediate expression of emotion, especially in the absence of external controls, and a more subjective and feeling-oriented approach to life. By contrast, the white group is more controlled, resistent to stress, and more pragmatically and realistically oriented.

It is not the purpose of this report to determine the reasons for these differences. They may be the result of cultural factors, genetic influences, motivational values, experience in test-taking, or any combination of these and still other forces. According to the results, however, these groups do differ in some important respects, and this differential response to the tests partially explains the observed differences in test validity for the two groups.

Our next steps in this report are first to investigate group differences on the performance (criterion) measures and second to relate test scores to actual performance in the most realistic and meaningful way possible.

RESPONSE OF WHITE AND NEGRO PATROLMEN TO THE PERFORMANCE (CRITERION) VARIABLES

The differences between the white and Negro subgroups for the eight criterion measures used in the study are shown in Table 13. The first three of these are regarded as the major criteria. The paired-comparison performance rating differs from the others in that it was specially instituted for this study and was not taken from records kept available by the Chicago Police Department.

The Three Major Criterion Variables

The white subgroup is rated higher (more favorably) on the paired-comparison performance rating, but for the Wave I sample the difference between the subgroups is not significant. However, this differences does reach statistical

CHAPTER VII -- Table 13

T-Tests for the Significance of the Differences between the Mean Scores for White and Negro Subgroups on the Performance (Criterion) Measures

Wave I Sample

		White Group			Negro Group				
Variable	Criterion	N	Mean	S.D.	N	Mean	S.D.	t	p
21	PC Rating	170	51.02	11.86	62	50.03	11.82	0.56	.580
372	1966 CPD Rating	170	85.72	3.34	61	84.67	3.98	1.83	.065
308	CPD Tenure	167	9.32	7.46	60	9.37	3.99	0.06	.954
373	CPD Awards	170	1.05	1.38	62	0.64	0.75	2.90	.004**(W)
375	IID Complaints	170	1.68	2.57	62	2.98	2.84	3.16	.002**(N)
374	Disciplinary Actions	170	.85	1.20	62	2.57	2.42	5.33	.000***(N)
359	Number of Arrests	170	40.42	28.74	62	80.66	60.54	4.99	.000***(N)
314	Times Absent	170	3.25	3.58	62	2.94	2.65	0.73	.525

Wave II Sample

		White Group			Negro Group				
Variable	Criterion	N	Mean	S.D.	N	Mean	S.D.	t	p
21	PC Rating	144	52.10	11.21	84	48.25	11.38	2.47	.014*(W)
372	1966 CPD Rating	184	86.39	3.91	90	83.58	4.11	5.37	.000***(W)
308	CPD Tenure	188	6.62	5.17	92	7.43	5.18	1.23	.218
373	CPD Awards	188	0.84	0.94	91	0.73	0.74	1.03	.305
375	IID Complaints	188	1.54	2.72	92	2.16	2.97	1.68	.091
374	Disciplinary Actions	188	3.46	4.28	91	7.27	6.98	4.76	.000***(N)
359	Number of Arrests	188	48.18	39.77	90	68.14	51.76	3.22	.002**(N)
314	Times Absent	187	2.48	2.47	89	3.20	3.33	1.81	.068

CHAPTER VII -- Table 13 cont.

Total Sample

Variable	Criterion	White Group N	White Group Mean	White Group S.D.	Negro Group N	Negro Group Mean	Negro Group S.D.	t	p
21	PC Rating	314	51.52	11.57	146	49.00	11.60	2.16	.029*(W)
372	1966 CPD Rating	354	86.07	3.66	151	84.02	4.09	5.29	.000***(W)
308	CPD Tenure	355	7.89	6.49	152	8.20	4.84	0.58	.568
373	CPD Awards	358	0.94	1.17	153	0.70	0.75	2.84	.005**(W)
375	IID Complaints	358	1.61	2.65	154	2.49	2.94	3.21	.002**(N)
374	Disciplinary Actions	358	2.22	3.46	153	5.37	6.06	6.00	.000***(N)
359	Number of Arrests	358	44.49	35.18	152	73.25	55.85	5.85	.000***(N)
314	Times Absent	357	2.85	3.08	151	3.09	3.07	0.82	.580

* $p < .05$
** $p < .01$
*** $p < .001$

(W) = White group mean higher
(N) = Negro group mean higher

significance for the Wave II sample (.02 level) and for the total sample (.03 level). Much the same pattern appears in the only other subjective supervisory rating used in the study--the one currently used and maintained by the Chicago Police Department--but at somewhat higher levels of significance. On this criterion, the difference between the white and Negro subgroups almost achieves significance for the Wave I sample (.07) and is significant beyond the .001 level of confidence both for the Wave II and for the total sample. It might be concluded from these supervisory rating figures that the white patrolman tends to be a better performer than his Negro counterpart. However, the possibility of supervisor bias in favor of the white patrolman must be considered. Such bias could easily cause greater differences between the subgroups on the CPD rating--which uses the traditional rating-scale techniques--than are found on the paired-comparison rating--which employs a technique generally recognized as reducing the effects of conscious or unconscious bias (Guilford, 1954, p. 174; Taylor & Wherry, 1951).

In contrast to these results, there are no significant differences between the subgroups for any of the samples on the tenure criterion. It seems safe to conclude that white and Negro patrolmen do not differ in departmental turnover rate.

The Six Remaining Criterion Variables

For these six variables, it is only on the absence record--as on the major criterion of tenure--that the subgroups show no significant differences. In general, the pattern for the remaining variables is that the white subgroup scores higher on the number of CPD awards made over their period of tenure and the Negro subgroup scores higher on the number of arrests made, the number of disciplinary actions taken against them, and the number of IID complaints sustained against them.

One feasible explanation for these differences is that Negro patrolmen tend more often to be assigned to predominantly Negro districts, which would include the ghetto areas. Since the crime rate in these districts is usually higher, it is natural that more arrests will be made. It also seems likely that patrolmen who are constantly "where the action is" will have more complaints made against them and, possibly, have more of these sustained.

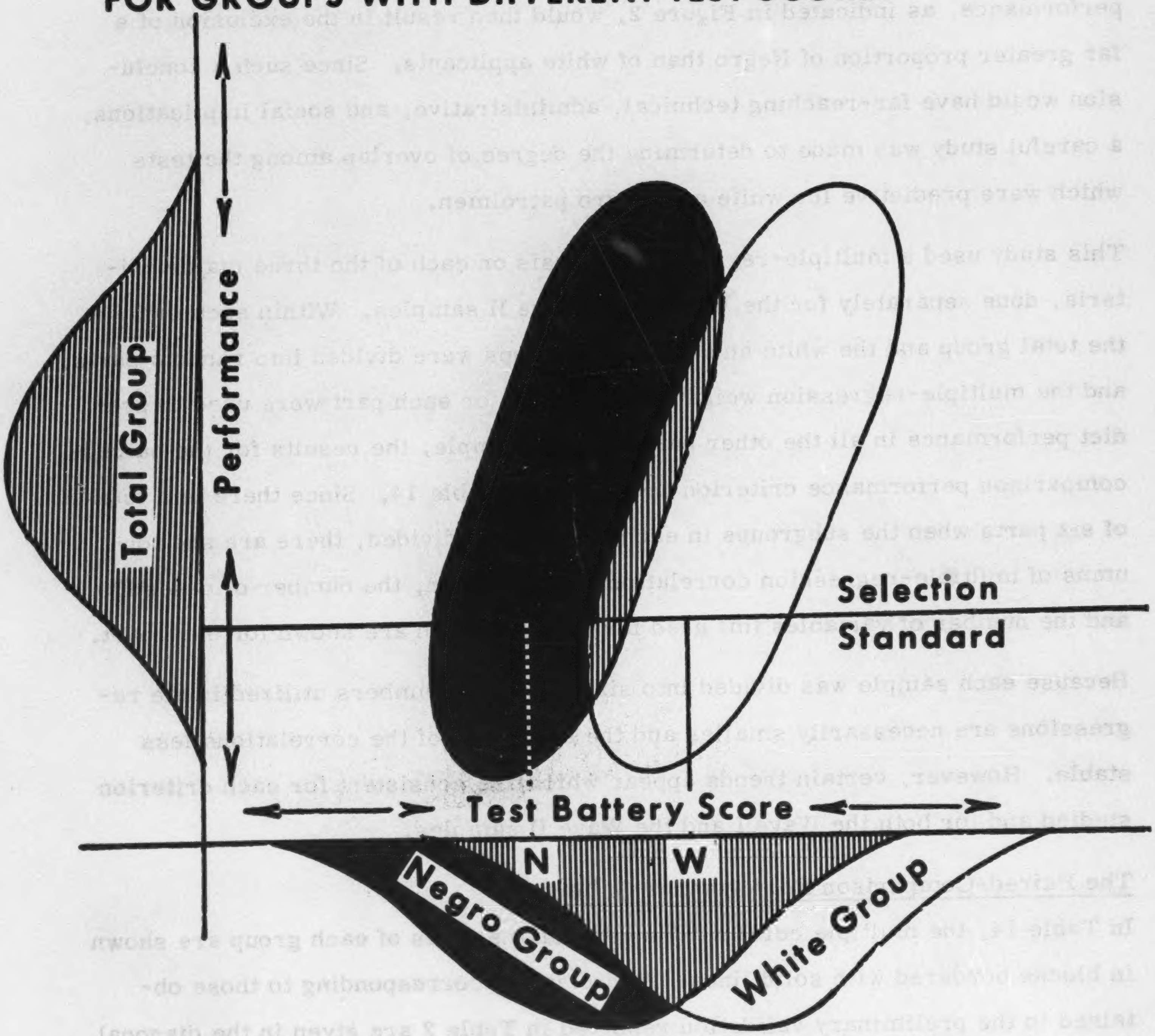
DIFFERENTIAL TEST VALIDITIES FOR WHITE AND NEGRO SUBGROUPS

The foregoing analyses have established that white and Negro patrolmen do respond differently on many of the predictor and criterion measures used in this study. This differential response almost certainly explains the significantly different test battery validities obtained.for the two groups as discussed in the multiple-regression-validation section of this Chapter and reported in Table 2.

The question now arises as to the most equitable and efficient way of predicting performance for this racially-mixed group of patrolmen. An analysis of this problem was pursued only for the three major performance or criterion variables, namely, the paired-comparison performance rating, the Chicago Police Department performance appraisal, and tenure. Various types of prediction model have been succinctly defined by Krug (1966). One of these, illustrated in Figure 1, might apply when there are no subgroup differences on some performance criterion but significant differences on certain test or predictor variables. In this study, tenure is the only major performance criterion where this model might apply. Clearly, if a particular selection standard is set with respect to such a criterion, the corresponding cut-off points on the predictor variables would be lower for Negro than for white applicants. Under these circumstances, use of different predictor-variable standards for the subgroups should not result in any change in the criterion variable for the total sample. Such a procedure is not an innovation in psychological measurement. In the past, different standards for specific predictor variables have often been applied to various subgroups in the population, such as those defined by age and sex. For example, different sex norms are frequently found in the manuals of psychological tests.

CHAPTER VII -- Figure 1

PREDICTION OF SIMILAR PERFORMANCE FOR GROUPS WITH DIFFERENT TEST SCORES


Total Group
Performance
Selection Standard
Test Battery Score
N
W
Negro Group
White Group

There is the possibility that the prediction of both supervisory ratings may conform to the model shown in Figure 2. However, it would be necessary to demonstrate that the criterion variables were free from bias and that the same tests predicted performance for both the white and the Negro subgroups. Under these circumstances, a lower test score could justifiably be regarded as indicating a lower level of performance. Setting a given selection standard for performance, as indicated in Figure 2, would then result in the exclusion of a far greater proportion of Negro than of white applicants. Since such a conclusion would have far-reaching technical, administrative, and social implications, a careful study was made to determine the degree of overlap among the tests which were predictive for white and Negro patrolmen.

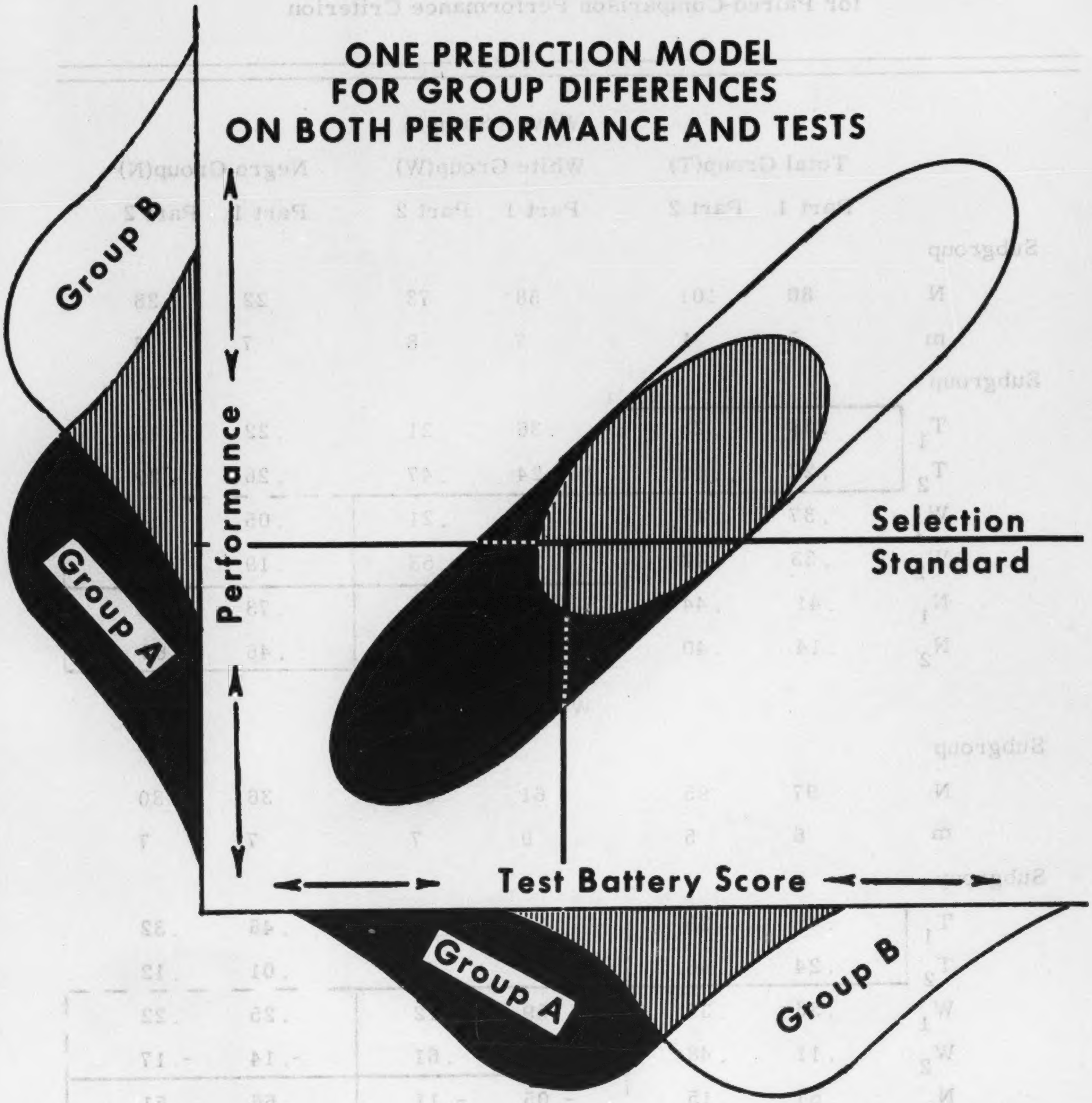
This study used a multiple-regression analysis on each of the three major criteria, done separately for the Wave I and Wave II samples. Within each sample, the total group and the white and Negro subgroups were divided into random parts, and the multiple-regression weights established for each part were used to predict performance in all the other parts. For example, the results for the paired-comparison performance criterion are shown in Table 14. Since there is a total of six parts when the subgroups in each sample are divided, there are six columns of multiple-regression correlations. In addition, the number of cases (N) and the number of variables (m) used for the prediction are shown for each part.

Because each sample was divided into six parts, the numbers utilized in the regressions are necessarily smaller and the estimates of the correlations less stable. However, certain trends appear which are consistent for each criterion studied and for both the Wave I and the Wave II samples.

The Paired-Comparison Performance Criterion

In Table 14, the multiple correlations between the parts of each group are shown in blocks bordered with solid lines. Correlations corresponding to those obtained in the preliminary validation reported in Table 2 are given in the diagonal entries in each block. Table 2 results are repeated here in that the primary

CHAPTER VII -- Figure 2



ONE PREDICTION MODEL FOR GROUP DIFFERENCES ON BOTH PERFORMANCE AND TESTS

Group B
Group A
Performance
Selection Standard
Test Battery Score
Group A
Group B

CHAPTER VII -- Table 14

Analysis of Within-Wave Multiple Correlations for Paired-Comparison Performance Criterion

	Wave I Sample					
	Total Group(T)		White Group(W)		Negro Group(N)	
	Part 1	Part 2	Part 1	Part 2	Part 1	Part 2
Subgroup						
N	80	101	58	73	22	28
m	5	4	7	8	7	6
Subgroup						
T_1	.38	.23	.36	.21	.22	.19
T_2	.27	.41	.24	.47	.26	.29
W_1	.37	.17	.59	.21	.05	.12
W_2	.33	.42	.21	.53	.19	.19
N_1	.41	.44	-.02	.25	.73	.36
N_2	.14	.40	.29	.37	.46	.66

	Wave II Sample					
Subgroup						
N	97	95	61	65	36	30
m	6	5	9	7	7	7
Subgroup						
T_1	.49	.25	.35	.05	.46	.32
T_2	.24	.46	.19	.40	.01	.12
W_1	.36	.33	.59	.12	.25	.22
W_2	.11	.48	.24	.61	-.14	-.17
N_1	.57	.15	-.05	-.11	.66	.51
N_2	.49	.46	.04	.08	.34	.75

validation coefficients for the racially mixed total group (.38, .41) are lower than those obtained for either racial group treated separately (white--.59, .53; Negro--.73, .66). These results are repeated for the Wave II sample, with moderate coefficients for the racially-mixed total group (.49, .46), increasing for the white group (.59, .66), and increasing even further for the Negro group (.66, .75).

The cross-validation coefficients between the parts of each group are shown in the off-diagonal entries in the solid-outline blocks. In both samples, the best cross-validations were obtained for the Negro group. In the Wave I sample, Negro-group primary coefficients of .73 and .66 shrink to .46 and .36. In the Wave II sample, Negro primary coefficients of .66 and .75 shrink to .34 and .51. The cross-validation coefficients for the white group show little improvement as compared with those for the total group but do, at least, remain relatively stable.

In contrast to these results, the poorest validity coefficients were obtained when predicting the performance of Negroes from weights established on white samples and vice versa. These coefficients are shown in blocks bordered by broken lines. All are at a low level, and some sink to zero or even become negative.

If the validity coefficients for either the white or the Negro group are based on weights obtained on the total racially-mixed group, they are inconsistent. Thus, the coefficient for one part of the white group is .33 as compared with .11 for the other part. Again, the coefficient for one part of the Negro group is .49 as compared with .15 for the other.

The Chicago-Police-Department-Rating Performance Criterion

Table 15 shows the results for this criterion. They are not quite as clear-cut as those obtained for the paired-comparison performance appraisal, but they follow the same trends for both the Wave I and the Wave II samples, as shown in the diagonals of the solid-line blocks. Thus, the primary validity coefficients

CHAPTER VII -- Table 15

Analysis of Within-Wave Multiple Correlations for C. P. D. Performance Rating Criterion

Wave I Sample

	Total Group(T)		White Group(W)		Negro Group(N)	
	Part 1	Part 2	Part 1	Part 2	Part 1	Part 2
Subgroup						
N	80	101	58	73	22	28
m	11	4	9	6	5	7
Subgroup						
T_1	.63	.20	.52	.24	.45	.07
T_2	.19	.53	.18	.39	.02	.37
W_1	.54	.10	.61	.21	.24	.01
W_2	.25	.53	.28	.54	-.01	.23
N_1	.67	.25	.38	.19	.65	.35
N_2	-.03	.52	-.10	-.03	.11	.71

Wave II Sample

	Total Group(T)		White Group(W)		Negro Group(N)	
	Part 1	Part 2	Part 1	Part 2	Part 1	Part 2
Subgroup						
N	97	95	61	65	36	30
m	6	5	10	5	11	6
Subgroup						
T_1	.54	.25	.39	-.07	.35	.24
T_2	.23	.44	.24	.25	.10	.29
W_1	.44	.33	.62	.15	.06	.08
W_2	.12	.38	.25	.39	.03	.10
N_1	.57	.17	.20	-.46	.80	.34
N_2	.37	.63	.20	-.05	.37	.74

for the racially mixed total group in the Wave I sample (.63, .53) remain at about the same level for the white group (.61, .54) but show a more perceptible increase for the Negro group (.65, .71). The same pattern holds for the Wave II sample, with an even more spectacular increase in the Negro-group coefficients (.80, .74).

The cross-validation coefficients between the parts of each group (shown in the off-diagonal entries in the solid-outline blocks) are generally lower for this criterion than for the paired-comparison appraisal. However, with the exception of a coefficient of .11 for Part 2 of the Negro group, the highest cross-validation coefficients are obtained when predicting the performance of Negro patrolmen from weights established on Negro groups.

For this criterion again, poor validity coefficients are obtained when attempting to predict performance across racial groups, i.e., when attempting to predict Negro performance from weights established on white groups or vice versa. Nine out of 16 of these coefficients are essentially zero, and one swings to negative .46.

Results of attempts to predict performance for white and Negro groups from weights based on a racially-mixed group are contradictory as they were for the paired-comparison criterion. Thus, coefficients obtained for the two white groups were .33 and .12, for the two Negro groups .37 and .17.

The Tenure Performance Criterion

Table 16 gives the results for this criterion. These present a contrast to those presented in the two previous tables in that all the coefficients, both primary and cross, are at generally high levels. However, the trend for cross-validity coefficients to be higher within racial groups and especially within the Negro groups is evident here too.

CHAPTER VII -- Table 16

Analysis of Within-Wave Multiple Correlations for Tenure Performance Criterion

Wave I Sample

	Total Group(T)		White Group(W)		Negro Group(N)	
	Part 1	Part 2	Part 1	Part 2	Part 1	Part 2
Subgroup						
N	80	101	58	73	22	28
m	11	12	7	14	6	8
Subgroup						
T_1	.69	.52	.67	.47	.53	.50
T_2	.57	.68	.51	.66	.57	.65
W_1	.72	.52	.70	.47	.51	.49
W_2	.50	.62	.45	.66	.49	.57
N_1	.70	.62	.62	.48	.80	.70
N_2	.77	.89	.69	.62	.83	.94

Wave II Sample

	Total Group(T)		White Group(W)		Negro Group(N)	
	Part 1	Part 2	Part 1	Part 2	Part 1	Part 2
Subgroup						
N	97	95	61	65	36	30
m	22	17	19	12	8	9
Subgroup						
T_1	.85	.75	.80	.75	.77	.66
T_2	.69	.77	.66	.74	.68	.57
W_1	.88	.77	.89	.79	.73	.63
W_2	.69	.80	.67	.87	.67	.44
N_1	.81	.71	.63	.68	.86	.71
N_2	.71	.70	.69	.59	.71	.81

SUMMARY AND CONCLUSION

This Chapter opened with an historical review of the events which led up to the heated public controversies of the past decade concerning the use of psychologically-based methods of personnel selection. A major concern in these controversies was that the use of tests might discriminate against disadvantaged groups in the labor market. This concern was given legal support and sanction by Title VII of the 1964 Civil Rights Act, which forbids discrimination by employers, labor unions, and employment agencies on the basis of race, color, religion, sex, or national origin. However, one clause in Section 703 (h) of Title VII, known as the Tower Amendment, was inserted to protect the right of employers to use tests for purposes of employment, placement, and promotion, provided their intention was not to practice discrimination as defined by the Act. As a further aid to employers, Title VII also established the Equal Employment Opportunity Commission whose primary initial task was to help employers in implementing both the letter and the spirit of the new law.

Following this historical review was a discussion of the various types of discrimination which can occur when tests are used. In this discussion, the principal focus was on the means of preventing such test-related discrimination, particularly on means recommended by the EEOC in its publications on this question. Those types of discrimination which can affect any member of the population can usually be eliminated if test administration, scoring, and interpretation are performed by people specially trained and, in the case of interpretation, professionally qualified for this work. Other types of discrimination, which are most likely to operate specifically against racial subgroups, can be prevented by using specialized techniques for the validation of tests and the implementation of selection programs. In this instance again, professionally trained personnel are essential.

Pressures from those Civil Rights groups which maintain that "all tests discriminate against the Negro," threats to complain to state Fair Employment Practices Commissions, and the seeming complexity of some of the requirements for guarding against discrimination have occasionally so intimidated employers as to cause them to abandon all use of tests. Some psychologists (e.g., Ash, March, 1966, p.11) have stated that this action is the equivalent of throwing out the baby with the bath water. Perhaps an even more apt analogy is that abandoning the use of tests is like throwing away a scalpel because it bared a malady. It would be ironic if the very instruments which first drew attention to the far-reaching effects of economic, social, and cultural deprivation were not now used to minimize these effects in the job market by diagnosing the exact training needs of an applicant so that he can be placed in an occupational situation where he is likely to succeed.

As a further irony, it is, of course, within the letter of the law for an organization indiscriminately to hire large numbers of individuals from so-called disadvantaged groups. The usual result of such indiscriminate hiring is that those hired either leave or are fired within a short time because they are unable to perform to their own or the company's satisfaction. This upshot is not only expensive and disruptive for the hiring organization but is also disastrous for the individuals in question. In the opinion of this investigator, a history of repeated occupational failure is the major cause of the attitudes of despair and resentment held by the disadvantaged seeking employment and of distrust and reluctance held by their potential employers. Proper selection can do much in the long run to ameliorate this situation. It is particularly important where the removal of an unsatisfactory employee is difficult after an initial probationary period has elapsed.

The results of the present investigation indicate that equitable and effective selection and placement are possible with multi-racial groups and that one method for achieving this goal, and probably one of the most effective, is to use specifically validated tests interpreted by professionally qualified personnel. Parenthetically, it should be noted that for effective interpretation of test results, the sex, age, and race of the individual must be known to the interpreter. Deliberate omission of this information can lead to discrimination against either sex in particular occupations, against the aged in many instances where speeded tests are used, and against the disadvantaged in most circumstances.

Specifically, our investigation has shown that the subgroups of white and Negro patrolmen included in the study differ significantly on certain of the test (predictor) variables and on six of the eight performance (criterion) variables. On the latter variables, many of the differences may be a result of disparate conditions for performance, such as assignment to districts, etc. However, significant differences on the test variables occur in all areas of the test battery, ranging from background data through work interests to intellectual abilities, special aptitudes, and behavior attributes. Whatever the origin of these differences (genetic, environmental, or a combination of these), they do exist and probably account for the differential test validities obtained.

The multiple regression analyses, done separately for the two samples in the study (Wave I and Wave II), both yielded higher validity coefficients for racially separate groups than for racially-mixed groups. Further, in predicting the two most significant performance criteria, the best cross-validation coefficients were obtained when the racial groups were treated separately, i.e., when weights based on Negro subjects were applied in predicting the performance of Negroes and vice versa. In contrast, the lowest cross-validation coefficients were obtained when predicting across racial groups, i.e., when weights based on Negro groups were applied in predicting the performance of white groups and vice versa. When weights derived from a racially-mixed group were used to predict the performance of either white or Negro groups, the results were inconsistent and contradictory. Interestingly enough, for each of the three major criterion measures, the highest validity and cross-validity coefficients were obtained for the Negro group.

On the basis of these results, separate validations for different racial groups are recommended as a routine procedure in the selection of patrolmen. Furthermore, this procedure should not be limited to the case of white and Negro groups but should be extended to any other minority groups, e.g., the Spanish-speaking, whenever there are a sizable number of such applicants. Under these conditions, the Police Department should be selecting the most suitable candidates from all races, and each candidate, regardless of race or national origin, will be provided with an equal opportunity to utilize his particular skills and abilities in achieving occupational success.

REFERENCES

Ash, Philip. Race, employment test, and equal opportunity. Industrial Management, March, 1966, 8, 8-12.

Ash, Philip. The implications of the Civil Rights Act of 1964 for psychological assessment in industry. American Psychologist, August, 1966, 21, 797-803.

Black, Hillel. They shall not pass. New York: Murrow, 1963.

Bragg, Emma W. A study of the in-put characteristics of the OMPER project trainees at Tuskegee Institute, Alabama. A report submitted to the officers of the OMPER project. Research, Experimentation and Demonstration Project, Contract No. 82-01-66-35, Project No. Ala. (M) 6082-000, June, 1967. (Multilithed report)

Campbell, J. The problem of cultural bias in selection: I. Background and literature. In The Executive Study Conference, Selecting and training Negroes for managerial positions. Princeton, N.J.: Educational Testing Service, 1965, 57-64.

Dreger, R.M. and Miller, K.S. Comparative psychological studies of Negroes and whites in the United States. Psychological Bulletin, 1960, 57, 361-402.

Dugan, Robert D. Current problems in test performance of job applicants: II. Personnel Psychology, Spring, 1966, 19, 18-24.

Equal Employment Opportunity Commission. Guidelines on employment testing procedures. Washington, D.C.: EEOC, August 24, 1966.

Gross, Martin L. The brain watchers. New York: Random House, 1962.

Guilford, J.P. Psychometric methods. (2nd ed.) New York: McGraw-Hill Book Company, Inc., 1954.

Guilford, J.P. Fundamental statistics in psychology and education. (3rd ed.) New York: McGraw-Hill Book Company, Inc., 1956.

Kirkpatrick, James J. *et al*. *Differential selection among applicants from different sociometric or ethnic background.* Final Report to The Ford Foundation, May, 1967. (Unpublished)

Klineberg, O. Negro-white differences in intelligence test performance: A new look at an old problem. *American Psychologist*, 1963, *18*, 198-203. .3

Krug, Robert E. Some suggested approaches for test development and measurement. *Personnel Psychology*, Spring, 1966, *19*, 24-35.

Lockwood, Howard C. Testing minority applicants for employment. *Personnel Journal*, July-August, 1965, *44*, 356ff.

Lockwood, Howard C. Critical problems in achieving equal employment opportunity. *Personnel Psychology*, Spring, 1966, *19*, 3-10.

Lopez, Felix M., Jr. Current problems in test performance of job applicants: I. *Personnel Psychology*, Spring, 1966, *19*, 10-18.

Lucas, C.M. Survey of the literature relating to the effects of cultural background on aptitude scores. *Research Bulletin, 15-13.* Princeton, N.J.: Educational Testing Service, 1953. (Multilithed report)

Mayfield, Harold. Equal employment opportunity: Should hiring standards be relaxed? *Personnel*, September-October, 1964, *41*, 8-17.

Standards for educational and psychological tests and manuals. Prepared by a joint committee of the American Psychological Association, American Educational Research Association, and The National Council on Measurement in Education (co-chairmen: John W. French and William B. Michael). Washington, D.C.: 1966.

Taylor, Erwin K. and Wherry, Robert J. A study of leniency in two rating systems. *Personnel Psychology*, Spring, 1951, *4*, 39-47.

Wallace, Phyllis, Kissinger, Beverly, and Reynolds, Betty. Testing of minority group applicants for employment. Equal Employment Opportunity Commission, Office of Research and Reports (*Research Report 1966-7*), March, 1966.

Whyte, William H., Jr. *The organization man*. New York: Simon and Schuster, 1956.

CHAPTER VIII

Specific Patterns of Patrolman Field Performance

An organization which makes diversified demands of its members and allows for individuality of style in meeting these demands will not be composed of behaviorally and psychologically homogeneous personnel. Rather, it will consist of a variety of subgroups, differing among themselves but marked by a general internal similarity in the performance of their members. A police department is such an organization. Therefore, it is reasonable to try to search out its various performance groups.

METHOD

In this study, such groups were located by a statistical method known as nodal analysis. This method was applied to each testing Wave separately, using the eight major criterion or performance measures described in Chapter III.

The specific technique of nodal analysis used followed the four steps suggested by Saunders and Schucman (1962). First, relationships between individuals are determined by squaring the differences of the scores of each separate pair of cases and adding these squares across the performance variables. In the second step, "clusters" are derived by combining, one at a time, those cases with the smallest differences. Thus, individuals are matched on the basis of similarity of performance patterns. The clustering procedure continues until no further clusters are grouped about that cluster's "node" or point representation, and mean scores are calculated for the criterion or performance variables. Finally, each case's distance from its node is calculated. The resulting clusters of patrolmen exhibit patterns of performance which are similar within each group (internally homogeneous) but which differ more markedly from group to group and from the general average of the patrolman population as a whole (externally distinct).

The mean performance scores for each node, identified by the nodal analysis specific to each testing Wave, were analyzed by means of a second nodal analysis to determine which performance styles were common to patrolmen in both Waves and which were unique to one Wave only. This procedure identified four subgroups with membership in both Waves. In a sense, the identification of matching nodal groups across the two Waves confirms the significance of that performance pattern. Two further nodal groups were found to have members unique to Wave I, and two others membership unique to Wave II. These unique groups have fewer members than the cross-Wave groups and may simply represent patterns of performance which are less common in the department.

Since the nodal analysis technique did not force every test case into a node, there remained a certain number of unassigned cases which could be considered only as representative of the total sample originally drawn from the 12 districts. As a check, the mean scores of this group were compared with those of the original sample and found to be equivalent. Hence, this group represents what might be called the "basic performance pattern" of the Patrol Division of the Department.

As an aid in interpreting the performance subgroups, the differences between the basic pattern group and each subgroup on the defining performance criteria and on all available psychological scores were tested for significance with t-tests. The discussion which follows considers only those variables which exhibited statistically significant differences equal to or better than the five per cent level of confidence. That is, the difference observed between the subgroup and the basic performance group could be expected, on the basis of chance, to occur only about five times out of 100.

DISCUSSION

The eight nodal subgroups identified in the analysis were thought to provide a significant basis for analyzing field performance and its relation to the psychological dimensions assessed by the tests. To facilitate the analysis, the subgroups were organized into a framework (Table 1) which clearly points up the differences in performance.

Our discussion will begin with the three patrolman subgroups which consist of relative newcomers to the police force, Subgroup 1 being rated high on performance, Subgroup 2 exhibiting average performance, and Subgroup 3 being rated low in performance. Next, three performance patterns characterizing patrolmen with average tenure will be discussed. Subgroup 4 exhibited desirable performance, while Subgroups 5 and 6 are discipline problems for the department. Finally, the two groups with longest average tenure will be examined. Here, the old-timers of Subgroup 7 appear to maintain a higher level of activity in the department than the patrolmen in Subgroup 8.

In the ensuing discussion, all subgroups are compared and contrasted to the "basic performance" group, which is probably representative of the average Chicago patrolman. The organization of the discussion by tenure (i. e., newcomers, established patrolmen, and old-timers) takes advantage of certain similarities and differences between the patterns of performance exhibited by the subgroups. While it might be desirable to contrast the test results of each subgroup with those of every other subgroup, time and space do not permit such a presentation.

Figures 1 through 3 contain the performance criterion profiles of the eight subgroups. The scores have been standardized and are presented on the

CHAPTER VIII -- Table 1

Eight Performance Pattern Subgroups

Successful Performance	Unsuccessful Performance
NEWCOMERS TO THE DEPARTMENT	
Subgroup 1 Low Tenure - Excellent Performance	
Subgroup 2 Low Tenure - Good Performance	
	Subgroup 3 Low Tenure - Poor Performance
ESTABLISHED PATROLMEN	
Subgroup 4 Average Tenure - Excellent Performance	
	Subgroup 5 Average Tenure - Poor Performance with Disciplinary Actions
	Subgroup 6 Average Tenure - Conflicting Ratings & General Disciplinary Problems
OLD-TIMERS	
Subgroup 7 Long Tenure - Excellent Performance	
Subgroup 8 Long Tenure - Good Performance	

equal-interval normalized standard score grid. Note that the average of the basic performance patrolman group is a standard score of 50 on each of the performance criteria. The performance of each subgroup may therefore be readily compared to that of the basic group. Table 2 gives the raw score means and standard deviations for all groups on the performance variables.

SUBGROUP 1

Newcomers with Excellent Performance

Criteria

The first subgroup consists of low-tenure officers who seem to be extremely active and effective. Although the majority of these officers are assigned to districts with an average index crime rate, their arrest and award records are in the very high range, significantly higher than those of the basic pattern group or of the other nodal groups. Their supervisors feel they are doing a good job as shown by their high-average ratings on both the paired-comparison and the department ratings, but they are not the highest-rated subgroup on these subjective measures. They show significantly fewer absences than the average, and an average level of disciplinary actions and complaints. Thus, Subgroup 1 certainly typifies a desirable pattern of patrolman performance.

Motivational Measures

Their background pattern tends to follow that of the basic pattern group with the two notable exceptions of Drive and Vocational Satisfaction as measured by the Personal History Index. Subgroup 1, scoring in the average range on these variables, is significantly higher than the basic group in both instances.

CHAPTER VIII -- Table 2

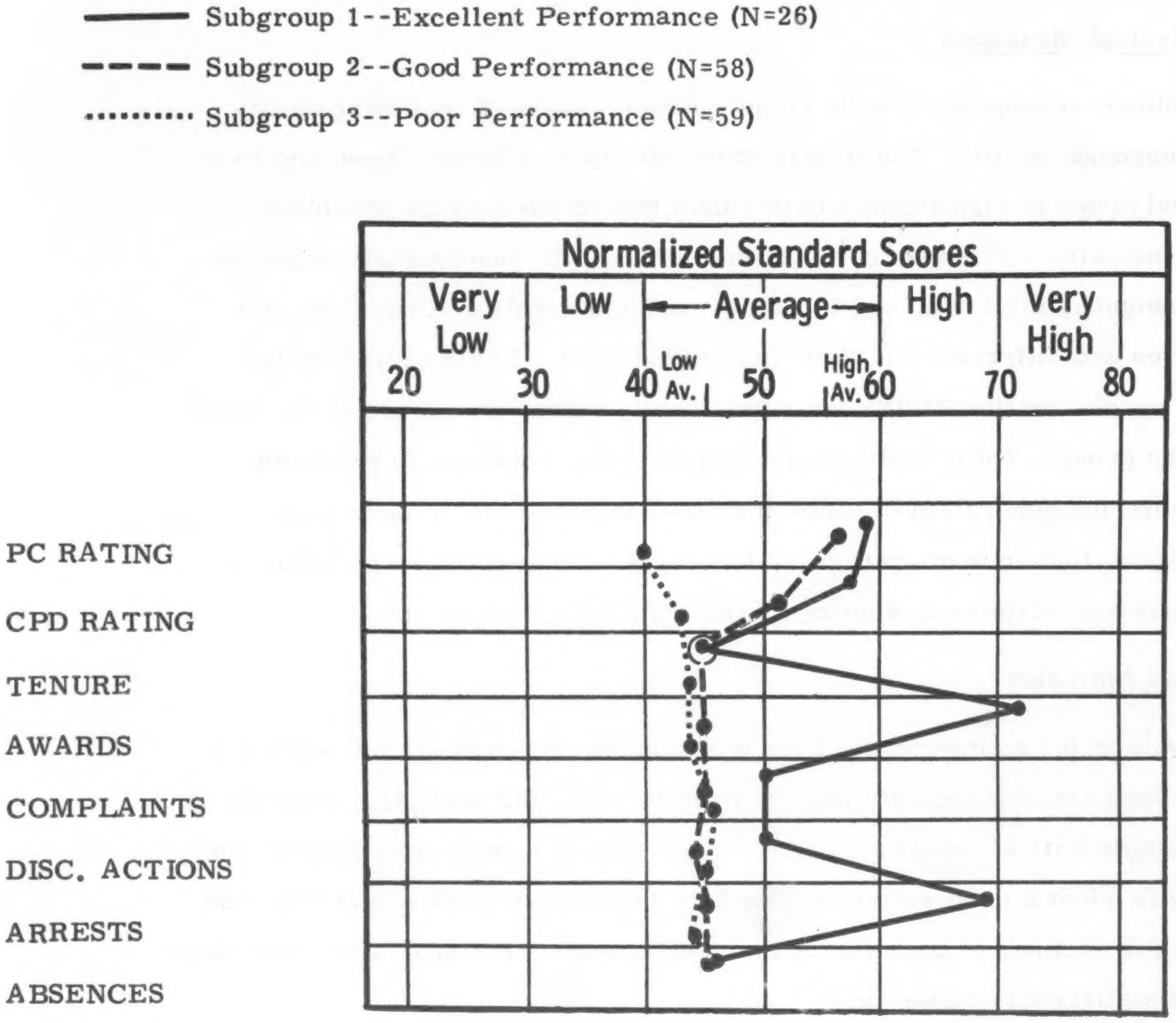
Raw Score Means and Standard Deviations for the Subgroups and
the Basic Performance Group on the Performance Measures

GROUP		Basic Performance	Newcomers			Established Patrolmen			Old-Timers	
			Subgroup 1	Subgroup 2	Subgroup 3	Subgroup 4	Subgroup 5	Subgroup 6	Subgroup 7	Subgroup 8
CRITERIA	N	202	26	58	59	35	24	18	11	13
P.C. Rating	Mean S.D.	48.60 10.82	61.24*** 8.36	56.37*** 8.07	39.74*** 5.05	63.94*** 4.02	41.44*** 8.60	52.16 10.01	63.68*** 5.17	60.68*** 10.30
Tenure (yrs.)	Mean S.D.	8.66 5.10	4.85*** 3.08	6.36** 4.64	4.54*** 4.12	10.11* 3.64	9.62 3.41	9.00 4.07	23.09** 17.26	16.15*** 5.27
Absences	Mean S.D.	3.23 2.56	2.23* 1.80	1.59*** 1.42	3.77 5.36	1.63*** 1.93	3.33 2.67	6.44** 4.57	3.91 2.31	3.08 2.06
Arrests (number)	Mean S.D.	56.87 44.43	105.88** 72.13	31.08*** 21.21	29.60*** 20.00	70.49 37.38	39.42* 30.76	87.67* 53.10	29.91** 26.05	79.38 47.72
CPD Rating	Mean S.D.	84.64 3.72	87.62*** 2.66	85.86** 2.71	82.68*** 3.10	90.19*** 2.87	82.62** 3.43	85.39 2.66	91.09*** 2.50	87.85** 3.49
Awards	Mean S.D.	0.72 0.67	3.50*** 1.90	0.34*** 0.40	0.26*** 0.35	1.20*** 0.74	0.57 0.69	1.26* 1.02	0.38** 0.36	1.67*** 0.76
Disc. Actions	Mean S.D.	3.54 5.07	2.69 3.34	0.71*** 1.19	1.24*** 1.62	1.36*** 1.71	5.38*** 1.78	13.10*** 7.59	0.32*** 0.72	7.51** 4.49
Complaints	Mean S.D.	2.43 2.71	1.52 2.22	0.31*** 0.93	0.63*** 1.07	1.06*** 1.79	2.19 0.98	7.87*** 5.46	0.73** 1.54	1.95 1.92

* The difference between this subgroup mean and the basic group mean could occur by chance 5 times per 100.
** The difference between this subgroup mean and the basic group mean could occur by chance 1 time per 100.
*** The difference between this subgroup mean and the basic group mean could occur by chance 1 time per 1000.

CHAPTER VIII -- Figure 1

Newcomers to the Department



The work interest responses of this subgroup suggest a stronger interest in artistic and stylized production, and a higher level of vocational aspiration.

Intellectual Measures

Patrolmen in Subgroup 1 have slightly higher levels of reasoning ability and language facility. Their perceptual ability on Closure Speed and Perceptual Speed is significantly better than that of the average patrolman. This indicates a stronger capacity on their part to meaningfully organize an ambiguous situation, and to rapidly and accurately discriminate similarities and differences in their perceptual field. Their scores on the AC Test of Creative Ability are significantly higher than those of the basic pattern group. Their ability to provide creative solutions to problems involving the generation of ideas and novel applications ranges from average to higher than average. This suggests that they are probably innovative and adaptive in solving problem situations on the job.

Special Aptitudes

Analysis of the scores on the Test of Social Insight shows that Subgroup 1 patrolmen are significantly less inclined to use withdrawal as a solution in a problematical social situation. They show a significantly greater tendency to adopt a cooperative approach in dealing with such situations and are more inclined to be active rather than passive, but this latter difference is not statistically significant.

Behavioral Measures

There were no significant differences between Subgroup 1 and the basic pattern patrolmen on the Arrow-Dot Test. On the Press Test, Subgroup 1 showed a significantly faster reaction time to both verbal and color stimuli, but their reaction under stress conditions (Part III), although somewhat faster, was not significantly different. In general, they showed a higher level of reactivity in all parts of the test.

Only one signficant difference was found in reviewing the performance of the two groups on three other behavioral tests. On the Hand Test, the Subgroup 1 patrolmen were significantly lower on Withdrawal. This suggests that their contact with reality is acceptable and satisfying.

The two patrolmen groups do not differ appreciably in creative potential as measured by the Cree Questionnaire, although Subgroup 1 patrolmen are significantly more conforming, and significantly less interested in scientific and theoretical work.

Subgroup 1 shows several significant differences on the behavior factors of the Temperament Comparator. As described in Chapter IV, these factors are composites based upon combinations of the 18 trait scores developed from the subject's self-description of his own behavior. Subgroup 1 officers are more outgoing (i. e., extroverted) and significantly less stable than the basic pattern patrolman group. Individuals scoring low on Stable are apt to exhibit a range of emotions and to be less predictable than their more stable counterparts. At the same time, such individuals are likely to be more active and more successful in creating enthusiasm in others. The significantly higher score of Subgroup 1 on the Excitable factor suggests that these men are lively and full of energy. However, they are significantly lower on Sociable, indicating that they are less socially inclined than patrolmen in general, and may be comfortable in solitary work not requiring many social contacts. This does not mean that they are anti-social, rather than they are more "work-oriented" than "person-oriented."

The Subgroup 1 patrolmen show significant departures from the basic pattern patrolmen in three of the 15 need areas measured by the Edwards Personal Preference Schedule. They are significantly lower on need Autonomy (to be independent, to feel free, to be able to come and go as desired), need Intraception (to analyze one's own and others' feelings and motives), and significantly higher on need Succorance (to receive support from others). These results suggest that they need direction and guidance in their work activity, are not overly concerned about, or sensitive to, interpersonal needs and functioning, and need some modicum of psychological support from other persons.

Summary

Subgroup 1 consists of younger patrolmen who react quickly yet correctly to situations presented to them. They are comfortable working in solitary situations, and, although disinclined to ponder the subtleties of human interaction, are enthusiastic, persuasive, and able to seek and obtain the cooperation of those they deal with. Rather than withdrawing from difficult situations, they are inclined to spontaneously meet and resolve them with solutions which are both creative and practical. In general, they accept the authority of those placed over them, and are effectively rewarded by the praise and encouragement of their supervisors. These men are highly rated and satisfied with their career choice. They exhibit the desire and the drive to achieve success and advancement. In spite of their low tenure and relative youth, they have exhibited a highly successful pattern of performance as patrolmen.

SUBGROUP 2

Newcomers with Good Performance

Criteria

Patrolmen assigned to the second subgroup by the nodal analysis procedure were significantly lower than the basic pattern patrolman group in length of service. Interestingly enough, in spite of the fact that they fall significantly below the average number of arrests and departmental awards, they were rated significantly higher than the basic patrolman group by their district supervisors in both the departmental and the paired-comparison ratings. Their attendance is significantly better than that of the basic group, and they received significantly fewer disciplinary actions and complaints. The high supervisory rating suggests that they are not guilty of inactivity. Therefore, the low number of arrests is probably not particularly undesirable in itself, although these men are assigned to districts with a slightly higher-than-average index crime rate.

Motivational Measures

The patrolmen in Subgroup 2 exhibit a pattern of background and experience scores which generally conforms to that of the basic pattern patrolman group, with three notable exceptions. First, they show a signficantly lower level of financial responsibility. This is no doubt related to their youth, but clearly they have not yet developed a high degree of skill in the management of their personal financial affairs. Another significant difference is in the area of parental family adjustment, which reflects the development of realistic and constructive attitudes towards the family, based upon early childhood experiences. Subgroup 2's score in this area suggests a better basis for emotional adjustment than that enjoyed by the basic group. Finally, this subgroup also exhibits a level of vocational satisfaction significantly higher than that of the basic group.

The occupational interests and aspirations of the members of Subgroup 2 do not differ significantly from those of the basic pattern patrolmen.

Intellectual Measures

This subgroup did not differ significantly from the basic group on reasoning ability, language facility, or visual perception. Their creative problem-solving ability is close to the average.

Special Aptitudes

The results of the Test of Social Insight suggest that this subgroup conforms to the average patrolman group in social orientation.

Behavioral Measures

In reviewing the results of the eight psychological tests which assessed the patrolmen's behavioral status, a suprising similarity was found between the members of this subgroup and the basic pattern patrolman group. Relatively few significant differences were identified. The reaction time of Subgroup 2 members was slightly higher but not significantly so. There were no differences between the two groups on the Arrow-Dot Test, the Maudsley Personality Inventory, or the Behavior Inventory. Three significant differences were found, however, on the Hand Test. Subgroup 2 patrolmen were less withdrawn (and hence in better contact with reality) and showed a significantly lower level of possible emotional pathology but a higher score on the Acting-Out ratio.

There were no significant differences on the Temperament Comparator, and the only difference on the Cree was a significant disinclination toward scientific and theoretical work situations on the part of the Subgroup 2 patrolmen. These patrolmen were also significantly lower on need Autonomy on the Edwards Personal Preference Schedule.

Summary

Although they are not outstanding on awards or number of arrests, the men included in Subgroup 2 are rendering good performance on the job. They are very similar to the basic pattern patrolman group in psychological make-up. Their childhood experiences have apparently laid the foundation for quite good adult emotional adjustment, as shown by the significantly lower pathology scores on the Hand Test. Their low level of financial experience might be considered a sign of personal irresponsibility and thus an area of concern, but since the group is characterized by significantly fewer complaints and disciplinary actions, their financial state apparently has not led to unethical or illegal behavior on their part. These men do not withdraw from reality, yet do not have any strong need for independence or autonomy. In all likelihood, they are able patrolmen who can function as part of a team, which may explain why their supervisory rating exceeds the objective performance measures of arrests and awards.

SUBGROUP 3

Newcomers with Poor Performance

Criteria

The third subgroup has, like the first, and second, a low level of tenure, but they do not appear to be enjoying the success of the groups already discussed. It is true that they have significantly fewer complaints and disciplinary actions against them, but inactivity rather than good performance appears to be the cause. The two supervisory rating categories place them significantly below average in performance. They made significantly fewer arrests than the basic pattern patrolman group, and received significantly fewer departmental awards. Their attendance is somewhat better than average, but they do not differ significantly from the basic pattern group in this respect. These patrolmen have the lowest average age of any of the subgroups (as suggested by their low tenure), and a majority of these officers are assigned to districts with higher-than-average crime rates.

Motivational Measures

Patrolmen in Subgroup 3 reveal a significantly lower level of family involvement and responsibility. They also show a significantly lower level of occupational or job stability. Low scores in these areas are partially a function of their relative youth; nevertheless, these scores suggest a disclination to undertake responsibilities, either of a family or job. These patrolmen are significantly higher than the basic pattern group on satisfaction with their vocational choice. However, such satisfaction is sometimes indicative of low aspiration and acceptance, rather than of a drive for achievement. The work interests of Subgroup 3 do not differ significantly from those of the basic pattern group.

Intellectual Measures

Subgroup 3 patrolmen are higher than the basic group in reasoning ability, and at the same level in the area of verbal comprehension. Their perceptual ability differs significantly from that of the basic group only in their Closure Speed performance. They show a higher level of ability in organizing a perceptual field. Their creative problem-solving ability does not differ from that of the basic group.

Special Aptitudes

Patrolmen in Subgroup 3 respond to problematical social situations in the same style as the basic patrolman group.

Behavioral Measures

Officers in Subgroup 3 exhibited a faster reaction speed and were significantly higher than the basic group in their reaction to color stimuli and in their reaction under the stress conditions produced in Part III of the Press Test.

These men score significantly higher than the average on Ego in the Arrow-Dot test, thought to be a measure of realistic control over impulsivity. They do not differ from the basic pattern patrolman group in the other areas of this test or in their responses to the Hand Test, Maudsley, and Behavior Inventories.

Their responses to the Cree Questionnaire suggest that they have a significantly lower level of creative potential and are inclined to a submissive social orientation. As to work orientation, they show a significant preference for unstructured work situations.

Although these men tend to be more outgoing and expressive in their behavior (Temperament Comparator, Factor I), they do not show any significant differences from the basic pattern group on this particular test.

Their results on the Edwards Personal Preference Schedule reveal no significant deviations from those of the basic patrolman group, except (perhaps) their significantly lower level of consistency. However, their responses are not so inconsistent as to be of questionable validity.

Summary

The patrolmen included in Subgroup 3 are relatively new to the job, but are already not well esteemed by their supervisors. They are close to the average patrolman in many respects, but their departures from this average provide insights into their psychological functioning and job performance. They are more satisfied than most patrolmen with their vocational choice, but it is apparent that good performance does not, in this instance, spring from such satisfaction. They show a lack of responsibility in the family situation and in occupational outlook. Their performance data suggest that they do not get into much trouble, and show good attendance, but do not do much on the street. Their mental endowment and aptitudes are average to above average. They have the ability to reason and to react quickly to situations, but apparently are not motivated to do so. They might be characterized as putting in what they consider the bare minimum of performance, a result perhaps of their lack of responsibility. In short, they have not been reached by their supervisors or by the intrinsic responsibilities of the job of police officer. These men have many acceptable qualities, and are well satisfied with their jobs, but this satisfaction is by no means shared by their supervisors. Many of these men are assigned to higher crime-rate districts, and, without speculating on whether or not their specific assignment may be one of the causes of their poor performance, it should be noted that these difficult districts are probably the last places they should be assigned.

SUBGROUP 4

Established Patrolmen with Excellent Performance

Criteria

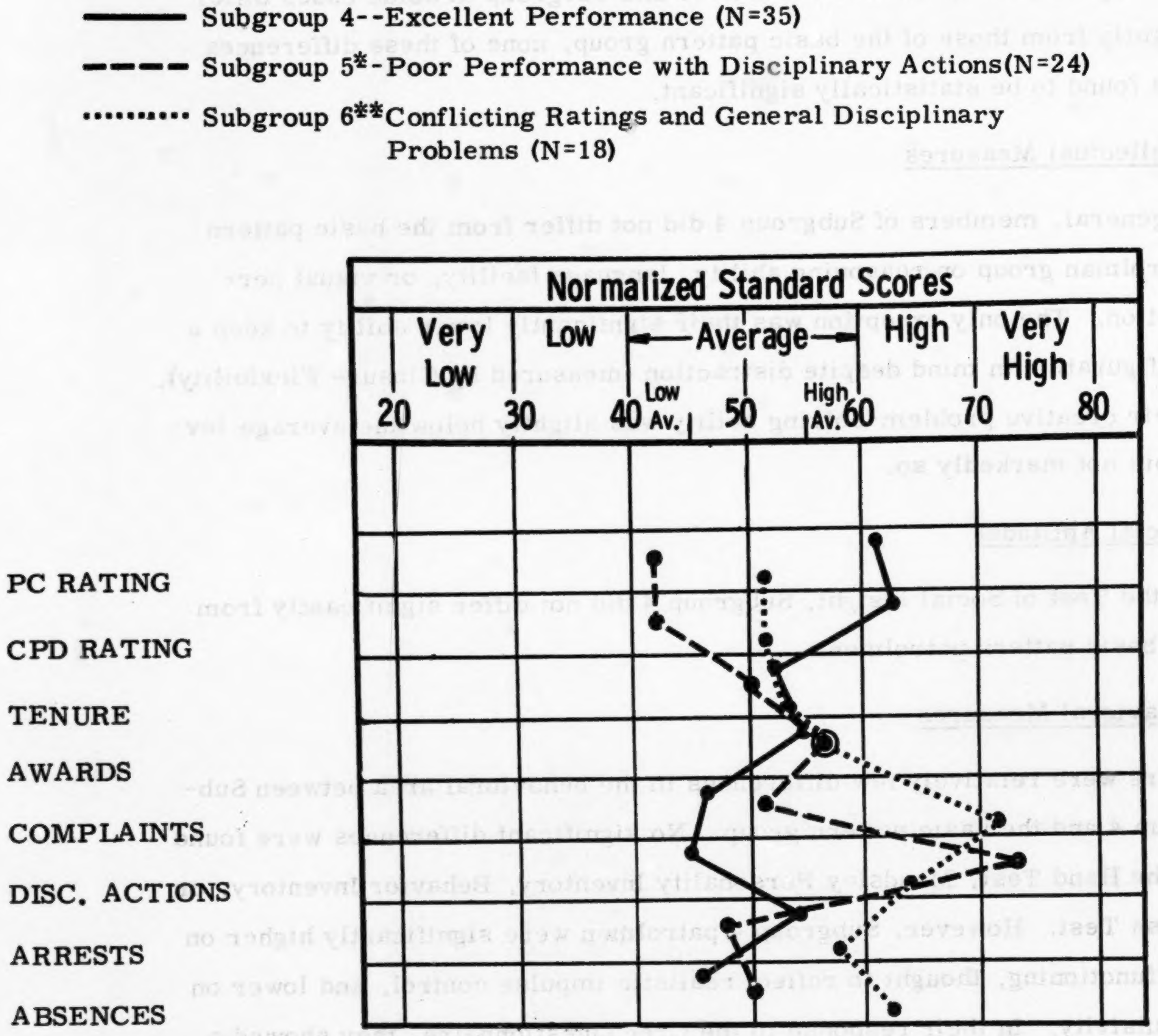
The patrolmen included in Subgroup 4 have significantly higher than average service tenure. Their performance is rated quite significantly above average by their district supervisors, and they have received an appreciable number of departmental awards. Their attendance is good, and they have received a significantly lower than average number of complaints and disciplinary actions. The number of arrests they made is also significantly higher than average. In general, this subgroup seems to represent a very desirable pattern of field performance. Officers in this subgroup are assigned to districts which represent the spectrum of law enforcement, i.e., high crime rate as well as lower-than-average crime rate areas.

Motivational Measures

On background and experience, Subgroup 4 differs on several points from the basic pattern patrolman group. They scored significantly lower on achievement in, and adjustment to, the school situation and also on special or technical accomplishment. Their scores in these two areas suggest that members of this group did not care very much for school, entered into marriage early, and did not pursue additional technical preparation or education. This is in part verified by a significantly higher score in the area of Early Family Responsibility, which reflects early assumption of family responsibilities and involvement in family activities. Subgroup 4 patrolmen also score higher than average in the area of occupational stability. This finding suggests that they are more concerned than the average patrolman with maintaining security and stability in their employment. They also had a significantly higher score on the General

CHAPTER VIII -- Figure 2

Established Patrolmen



* Subgroup is unique to Wave I.
** Subgroup is unique to Wave II.

Health factor, which points to a better than average freedom from physical ailments and from lost work time resulting from illness.

Although the occupational interests of this subgroup in some cases differ slightly from those of the basic pattern group, none of these differences was found to be statistically significant.

Intellectual Measures

In general, members of Subgroup 4 did not differ from the basic pattern patrolman group on reasoning ability, language facility, or visual perception. The only exception was their signficantly lower ability to keep a configuration in mind despite distraction (measured by Closure Flexibility). Their creative problem-solving ability was slightly below the average level but not markedly so.

Special Aptitudes

On the Test of Social Insight, Subgroup 4 did not differ significantly from the basic pattern patrolmen.

Behavioral Measures

There were relatively few differences in the behavioral area between Subgroup 4 and the basic pattern group. No significant differences were found on the Hand Test, Maudsley Personality Inventory, Behavior Inventory, or Press Test. However, Subgroup 4 patrolmen were significantly higher on ego functioning, thought to reflect realistic impulse control, and lower on impulsivity. In their response to the Cree Questionnaire, they showed a considerably stronger preference for involvement in social activities. Their Temperament Comparator results suggested that they are significantly less demonstrative and somewhat more talkative than the basic pattern group. They also showed a significantly lower level need Autonomy, as measured by the Edwards Personal Preference Schedule, and a higher level of need Aggression.

Summary

The patrolmen who consitute Subgroup 4 exhibit a good to excellent performance pattern, as indicated by both the objective and the subjective performance criteria. These men had a low level of academic accomplishment and an early assumption of family responsibilities with close involvement in family activities. They enjoy good health, and have a realistic control of their impulses and little inclination to demonstrate their feelings. They are socially involved and talkative, and are probably quite conventional and concerned with meeting the expectations of others. Their aggressive needs are significantly higher than those of the basic pattern patrolmen, but these needs appear to be constructively channeled. Their low scores on complaints and disciplinary actions support the validity of this interpretation. This group seems to represent middle-aged patrolmen, interested in the security of their job, who more than meet the expectations of their supervisors. Because of their poor academic preparation, they may have difficulty rising in the organization beyond their current status. However, they are not likely to leave the department for another position, and therefore in many ways represent a quite desirable component in the manpower of the organization.

SUBGROUP 5

Established Patrolmen with Poor Performance and Disciplinary Actions

Criteria

The performance record of this subgroup is clearly marginal compared with that of the basic pattern group. Members of Subgroup 5 were rated significantly lower in performance by their district supervisors in both the paired-comparison and the departmental ratings. They show an exceptionally high number of disciplinary actions. At the same time, their productivity, as expressed in number of arrests, falls significantly below that of the basic group. They show slightly more complaints and slightly fewer awards. Their tenure and attendance record fall within the average range. Obviously, this group's members are held in low esteem by their supervisors, and represent an undesirable pattern of field performance. Patrolmen in Subgroup 5 are assigned to high crime rate as well as lower-than-average crime rate districts.

Motivational Measures

On background and experience, there is only one area, vocational stability, in which this subgroup differs significantly from the basic pattern group. They are concerned with establishing security and stability in the work situation and with maintaining what has been achieved. Members of this group scored slightly lower in their drive to achieve success and advancement and slightly higher on parental family adjustment. Since this subgroup was specific to Wave I, no data on occupational interests were available.

Intellectual Measures

The patrolmen in Subgroup 5 did not differ from the average on reasoning ability, but were significantly lower on verbal comprehension (i.e., the ability to read and comprehend written material). Their scores in the perceptual area were generally slightly below average, and they were significantly below average in their ability to rapidly organize a set of apparently unrelated elements into a meaningful whole. However, these patrolmen were not significantly different in creative problem-solving ability.

Special Aptitudes

On the Test of Social Insight, this group showed a slightly less cooperative and slightly more passive orientation, but none of the differences was significant.

Behavioral Measures

The Press Test responses of this subgroup tended toward a slower reaction time than the average, but this tendency was not significant. Their scores on the Arrow-Dot test were close to the average. However, it is interesting to note that the group exhibits a profile pattern of impulsivity identified by the Employee Appraisal Section of the Personnel Division and thought to be associated with poor performance.

The group's scores on the Hand Test were not significantly different from those of the basic pattern group, but there was a tendency to score above the average in the area of withdrawal and pathological responses. The same situation obtains in the responses to the Maudsley Personality Inventory. There were no significant differences but a tendency to score higher in neuroticism.

The patrolmen in this subgroup scored significantly higher than average on the Behavior Inventory, an instrument specifically validated against a criterion of pass/fail on the personal background investigation conducted by the department.

The Temperament Comparator scores of Subgroup 5 did not differ markedly from those of the basic pattern patrolman group, although this subgroup tended to be somewhat more outgoing and sociable but somewhat less self-confident.

Summary

The patrolmen in Subgroup 5 exhibit a pattern of performance which is decidedly marginal. Their background scores indicate a concern for maintaining the status quo, and a corresponding lack of interest in self-development. Their tendency to score lower on achievement drive may be related to this. They are weak in one area essential to patrolman performance--the ability to understand written material. They also demonstrate a significantly lower level of ability to rapidly and correctly organize their perceptual field. The trends of the behavioral tests suggest adjustment problems, but the only clear-cut problem indicator was found in the results from the Behavior Inventory. These patrolmen report positive attitudes concerning their early developmental experiences, which should provide a good basis for emotional adjustment. Their tendency toward maladjustment presents an interesting contradiction to this premise. Apparently these patrolmen are marginal in both the language and the behavioral areas. They may in fact be "false positives," i. e., individuals who erroneously passed the screening of the Civil Service Commission and the department.

SUBGROUP 6

Established Patrolmen with Conflicting Performance Ratings and General Disciplinary Problems

Criteria

Patrolmen in Subgroup 6 have average tenure and receive average performance ratings from their supervisors. However, on the remaining performance criteria they present an interesting contrast. They are significantly above the basic pattern patrolman group in numbers of arrests made and departmental awards, but are even more significantly above the average on absenteeism, disciplinary actions, and complaints. Apparently the good work done by these men is overshadowed by the problems they cause the public and the department. They seem to attempt to do what is expected of them, as indicated by the arrests and awards criteria, but are in some way inept in their performance, in that complaints and disciplinary actions follow. Their attendance record suggests that they are less than responsible about reporting for duty. This subgroup is the only one which differs significantly from the basic patrolman performance group in the area of rater agreement. The supervisors who provided the paired-comparison performance ratings on these patrolmen were found to show lower overall agreement on the performance of these men. In general, this pattern of performance seems undesirable, since the field supervisors do not recognize or else disagree on the true caliber of these officers, and may give them assignments beyond their competence, which they are therefore unable to handle. Officers in this subgroup are not found in any one type of district, but rather are assigned to high crime rate, average, and low crime rate districts.

Motivational Measures

The background and experience scores of Subgroup 6 patrolmen are quite similar to those of the basic pattern patrolman group. The only exception is on Early Family Responsibility, where Subgroup 6 scores lower. This suggests that many of these men have not yet assumed family responsibilities (even though their average age is 35) and are not involved in the activities of their own immediate families.

In the main the work interests of this group closely parallel those of the basic pattern patrolman group. The only exception is their significantly greater preference for work situations involving authority over others and prestige.

Intellectual Measures

The patrolmen in Subgroup 6 tend to score below the basic pattern group in the perceptual area, and were significantly lower on Closure Speed. This suggests an inferior ability in rapidly organizing an apparently unrelated set of elements into a meaningful whole.

Special Aptitudes

Although no statistically significant differences exist between this group and the basic pattern, members of Subgroup 6 exhibit an interesting tendency to utilize withdrawal as a means of coping with problematical social situations.

Behavioral Measures*

The Subgroup 6 patrolmen tend to react slightly faster than members of the basic pattern group. Furthermore, as measured by the Arrow-Dot Test, they exhibit contrasting, equally strong levels of impulsivity and of moralistic control of impulse. This mode of adjustment suggests that a strong conflict exists between what these men may want to do and what they know is the ideal thing to do. They do not achieve a realistic resolution to this conflict, but rather alternate between the two extremes.

Their responses to the Cree Questionnaire show a significantly stronger orientation toward systematic and prescribed activity on the job. They also describe themselves as having a significantly faster level of reaction speed.

In the temperament area, the most noteworthy finding is the significantly lower score shown by these men on self-reliance. This suggests that they find it difficult to initiate and enforce decisions, are influenced by the wishes and desires of others, and are in need of social support. They are significantly less even-tempered than the basic pattern patrolmen and significantly more talkative. However, on the Edwards Personal Preference Schedule their scores conform to the average.

Summary

This subgroup's results on the performance criteria present an interesting contradiction. Subgroup 6 patrolmen show quite good performance in the areas of number of arrests and awards, and exceptionally poor performance in the areas of departmental disciplinary actions, IID complaints, and attendance. Essentially they are undependable and unpredictable, sometimes quite good, sometimes quite bad. This erratic performance is the probable cause of the relatively lower agreement shown by field supervisors in rating these patrolmen. Some supervisors may have been exposed to inadequate performance while others may not. Several points in the test data may account for this erratic performance. These men have not yet assumed family responsibilities, and may have entered police work for the authority and prestige it could bring them. Their low level of self-reliance suggests a need for some sort of ego support, and the badge and revolver may fill this need. They were below average in their ability to cope effectively with the ambiguity in the Closure Speed Test, and they were prone to use withdrawal in the Social Insight Test. These results suggest an inability to deal effectively and realistically with concrete problem situations. This inability is also reflected in the results on the Arrow-Dot Test, where there is a conflict between impulsivity and moral control. These patrolmen do not seem to possess the "street sense" of a good police officer in dealing with job problems. They may at times enter a situation impulsively, and at other times needlessly withdraw when they are needed. This behavior is probably rooted in their lack of self-confidence and responsibility, in their inability to realistically handle the problems which constantly confront a police officer, and in their own lack of impulse control. Because they are able to function acceptably for a certain percentage of the time, their performance ratings emerge as "average." From a broader perspective, they are by no means average and should be recognized as the marginal patrolmen they actually are.

---

*This subgroup was specific to Wave II. Hence, only Wave II test battery results are available for discussion.

SUBGROUP 7

Old-Timers
with Excellent Performance

Criteria

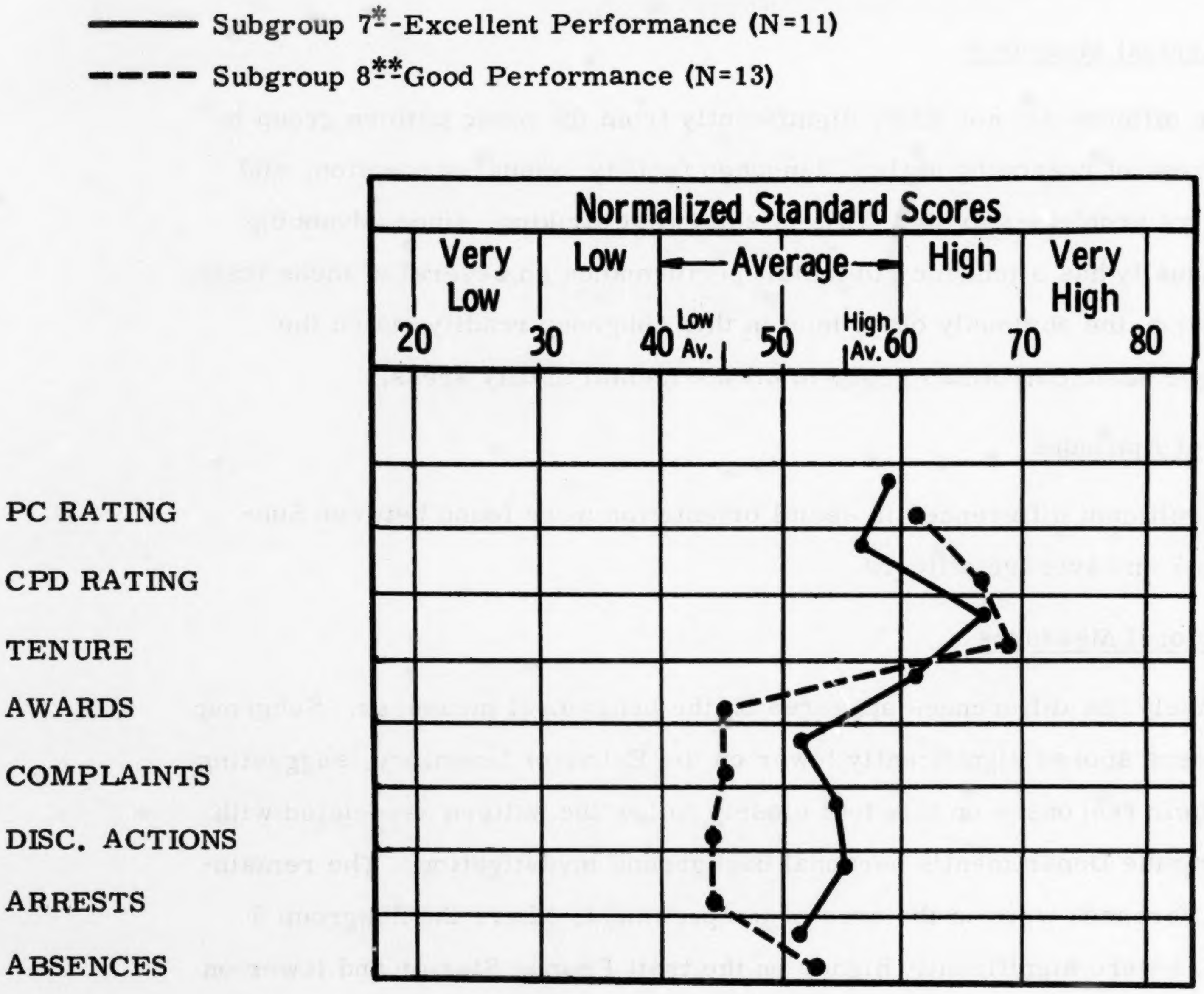
Patrolmen assigned to Subgroup 7 have quite significantly higher tenure than the basic group patrolmen, the highest tenure of any of the nodal subgroups. Their paired-comparison and departmental supervisory ratings are also considerably above average. Although these men are significantly lower in disciplinary actions and complaints, they are also significantly lower in number of arrests and in departmental awards. Their attendance record is average. This subgroup is composed of older patrolmen who may not be as active in the field as they once were (their paired-comparison or field rating is lower than their departmental or total one), but who are nonetheless considered to be excellent patrolmen by their supervisors. They are not highly productive in the sense of making many arrests or earning many awards, but are thought to provide good performance on the job. They are assigned to districts representing the entire range of index crime rate.

Motivational Measures

Officers in Subgroup 7 score significantly higher than the basic pattern patrolman in the area of Higher Educational Achievement. This suggests more education than the basic group and a later vocational start, combined with an interest in professional development. These men also exhibit a significantly high level of Financial Responsibility, reflecting mature management of their own personal resources. Furthermore, they score significantly higher on early assumption of family responsibilities and involvement in family acitivities. They have now established occupational stability, and are more concerned with the maintenance of what they have already attained than with plans for further development. Their level of Vocational

CHAPTER VIII -- Figure 3

Old-Timers



* Subgroup is unique to Wave I.
** Subgroup is unique to Wave II.

Decisiveness is significantly below that of the basic pattern patrolmen, indicating that they did not make a decisive occupational choice early in life but rather had a late start in their police careers.

Intellectual Measures

These officers did not differ significantly from the basic pattern group in the areas of reasoning ability, language facility, visual perception, and creative problem-solving. This is somewhat striking, since advancing age usually has a tendency to inhibit performance on several of these tests. However, the obviously older men in this subgroup readily match the younger basic patrolman group in all the mental ability areas.

Special Aptitudes

No significant differences in social orientation were found between Subgroup 7 and average officers.

Behavioral Measures

Relatively few differences appeared on the behavioral measures. Subgroup 7 officers scored significantly lower on the Behavior Inventory, suggesting that their responses on this test closely follow the pattern associated with passing the Department's personal background investigation. The remaining differences were in the area of temperament, where the Subgroup 7 officers were significantly higher on the trait Prompt Starter and lower on the trait Socially at Ease and on Sociable (Temperament Comparator-Factor V). These findings suggest a "work" rather than a "person" orientation in their day-to-day activities.

Summary

The performance criteria suggest that patrolmen in this subgroup are less active in the field than basic pattern patrolmen (a result, perhaps, of their age) yet maintain the high esteem of their supervisors, probably a consequence of many years of good service. These men assumed family responsibilities early and established good financial management of their affairs. This mature behavior was also demonstrated in their continuation of technical or professional training after completing academic studies. Although Subgroup 7 patrolmen did not show a strong early commitment to a law enforcement career, they have nevertheless established themselves securely, with a record of good performance over the years. These officers show a good level of mental endowment, especially when one remembers that they are much older than the basic pattern patrolman group. They do not appear to differ very much from this latter group in their aptitude or behavioral characteristics, but are clearly differentiated by the Behavioral Inventory screening procedure, and show a "work" orientation in their temperament. Despite their long tenure, these patrolmen have remained at a high level of effectiveness, and are thought to represent a successful pattern of patrolman performance.

SUBGROUP 8

Old-Timers
with Good Performance

Criteria

The patrolmen in this subgroup have a quite high level of tenure. Both of their supervisory ratings are significantly higher than average. They are significantly higher than the basic pattern men on awards but also on disciplinary actions. Their attendance, arrest, and complaint records were within the average range. Except for the disciplinary actions, they appear to be a quite successful patrolman group. Although it might be expected that older patrolmen would tend to be found working in slow crime rate areas, these men are assigned to both high and low crime rate districts.

Motivational Measures

On motivation, Subgroup 8 patrolmen exhibited a number of significant departures from the average of the basic pattern group. They showed a marked disinclination to participate in group activities, as evidenced by a significantly lower score in the area of School Activities (reflecting group participation in school) and in Leadership & Group Participation (reflecting group participation in adulthood). As a group, they show a significantly lower level of selling experience in their past employment. However, they showed a signficantly higher level of vocational stability, suggesting more concern with establishing security in the work situation than with improvement or development. Yet the patrolmen in this group indicate a signficantly lower level of satisfaction with their vocational choice.

The work interest profile of Subgroup 8 does not differ from that of the basic pattern group, except that their level of socio-economic aspiration is significantly lower.

Intellectual Measures

These men had lower scores in the perceptual area and significantly lower ones on the tests of Closure Speed and Closure Flexibility included in this area. Since these patrolmen are a good deal older than the average, their intellectual functioning may be somewhat hampered by deterioration resulting from advancing age.

Special Aptitudes

The results of the Test of Social Insight suggest one important difference in social orientation from the basic patrolman group. Subgroup 8 patrolmen were significantly less apt to utilize an aggressive approach to resolve the problematical social situations described in the test.

Behavioral Measures

There were no significant differences on either the Press Test or the Arrow-Dot Test. However, Subgroup 8 patrolmen scored significantly lower on the total score for the Cree Questionnaire, indicating a lower potential for creative work and a greater degree of conformity than shown by the basic pattern patrolmen. In the area of temperament, these men were significantly less sociable than the basic group, that is, they are probably less concerned with seeking out and enjoying the company of others. A number of significant differences were identified on the Edwards Personal Preference Schedule. Subgroup 8 patrolmen were significantly higher on need Order, need Affiliation, and need Nurturance, and significantly lower on need Autonomy. This suggests that they have to be told what to do by others, must perform in a precise and organized way, and are cooperative and loyal and enjoy assisting the unexperienced or troubled.

Summary

The patrolmen in Subgroup 8 present a pattern of good performance, with the exception of a higher-than-expected level of disciplinary actions.

Three areas in the Personal History Index support a finding in the temperament area which suggests that these officers are socially aloof and not inclined to seek out and enjoy the company of other persons. While they are quite concerned with maintaining job stability and security, they are dissatisfied with their occupational choice and pessimistic about improving their socio-economic level. This pessimism may account for the need for disciplinary actions in a group that is providing an otherwise acceptable level of performance. Age may play an underlying role in the above findings, as well as in the relatively lower level of mental prowess exhibited by the subgroup and in a disinclination to use aggression in solving difficult social situations. The members of the group may in part reflect the stereotype of the old-time beat patrolman, who was told specifically what to do and expected to carry out orders to the letter, who was loyal to the organization while at the same time giving aid to the underdog. The curious twist is the social disinvolvement of these patrolmen, which may be a result of constant exposure to the worst side of human nature.

CONCLUSION

The aim of this chapter was to identify and provide a description of meaningful performance subgroups contained within the patrolman population assessed and tested. The performance criteria scores had been subjected to a form of multivariate analysis which in essence measured the differences in the scores of each pair of individuals in the sample, and then assigned these individuals to subgroups on the basis of similarity of performance pattern. An extensive discussion of the differences in job performance and test results contrasted the basic performance pattern patrolman subgroup (the "average" patrolman, whose performance on the eight criteria was within average limits) with the eight emergent subgroups (whose performance pattern was distinctively different from the average).

The aim of this phase of the study is felt to have been accomplished, since subgroups were identified and since the psychological tests did provide adequate description and at least minimal differentiation from the average. This is not to say that we believe we have identified all possible performance subgroups. All of the subgroups discussed in this chapter were thought to be significant, but the utilization of additional performance criteria might very well result in the identification of additional meaningful subgroups. It is hoped that the subgroups discussed in this chapter will provide insight into the psychological make-up of the patrolman, and serve as strong evidence to support the hypothesis that over and above general requirements for suitability, there will be identifiable areas of individual difference between patrolmen which determine their success or failure on the various measures of performance in the field.

The bulk of our discussion has been concerned with contrasting the subgroups with the basic patrolman group. It was thought that this contrast would provide maximum definition of the psychological functioning of each type, and facilitate understanding and maximum utilization of the results. The value of this work is to be realized in the early identification of patrolmen who resemble members of a given subgroup, and in the proper placement (or rejection) of these individuals. Early identification may be important, since proper training and development of Subgroup 1 patrolmen, for example, would qualify these men for the advancement they obviously have merited by their performance. Proper placement is another question, and the presence of Subgroup 3 patrolmen in high crime rate areas is given as an example of apparent misplacement.

The specific manner in which these results may aid a particular department is an open question. A number of techniques are available, but immediate generalization of the results obtained in Chicago to other departments would seem to be premature. Continuing study of the significance of these groups and of the effects of factors not discussed in this chapter (e.g., effect of training, specific assignment, race, etc.) is indicated. The areas of psychological functioning in which the subgroup differs from the basic patrolman group would seem to provide a first step in identification for purposes of decision-making. The next stage of analysis would be to contrast the scores of each subgroup with those of every other subgroup, in order to determine the specific differences between each pair of groups with respect to all test variables. This procedure would provide a direct answer to the question of how to differentiate among subgroups, for example, Subgroups 1, 2, and 3. The results of such an analysis could be used to determine the psychological test variables to be used in a conventional multiple discriminant analysis for assigning patrolmen to the eight emergent subgroups.

We feel it desirable to conclude this chapter with some summary statement on the tests which seem most adequate in differentiating between performance subgroups. The specific differences discussed throughout the chapter provide the best answer to this question, but a listing of the tests thought to have the greatest likelihood of differentiating between the subgroups is presented in Table 3. These tests were identified through the use of an analysis of variance design, and, on the basis of present information, recommended for inclusion in further research on the performance subgroups aimed at the question of job placement and assignment.

REFERENCES

Saunders, D. R. and Schucman, H. Syndrome analysis: An efficient procedure for isolating meaningful subgroups in a non-random sample of a population. A paper read at the third annual meeting of the Psychonomic Society in St. Louis, September 1, 1962.

CHAPTER VIII - Table 3

Tests Identified by Analysis of Variance as Discriminatory Among Performance Subgroups

AREA OF MEASUREMENT	TEST
Motivation	Personal History Index (especially factors 2, 5, 6, 8, and 13)
Intellectual Ability	Closure Speed Closure Flexibility Perceptual Speed
Behavior	Press Test Temperament Comparator Arrow-Dot Test Behavior Inventory Edwards Personal Preference Schedule

CHAPTER IX

Summary of Results for Psychological Assessment of Patrolman Qualifications in Relation to Field Performance

GOALS OF THE STUDY

The issue of law and order vs. crime in the streets is likely to become one of the hallmarks of the late 1960's. Given the importance of the law-enforcement function, the importance of the individual police officer performing that function is patently obvious. Few occupations in recent years have been the object of so much public attention, debate, and controversy as that of the police officer. Police organizations have made impressive technical advances in such areas as communications systems and non-lethal weaponry but surprisingly few advances in the area of personnel selection. Since the individual police officer remains the single most important law-enforcement representative in society, it is unfortunate that contributions to his function made by the physical sciences have not been matched by the behavioral sciences. In the long run, we feel that better police officers are of greater importance than police tools for improved law enforcement.

The value of psychological testing in selecting police officers has long been recognized. However, the necessity for empirical validation of the tests has by no means met with the same recognition. We therefore thought that there was an essential need for a comprehensive research study in the police selection area. Three basic requirements had to exist before such a study could be initiated:

1. A group of behavioral scientists interested in addressing the problem,
2. The availability of financial support, and, perhaps most important,
3. A law-enforcement organization at a stage of development which would permit scientific access and freedom of inquiry while at the same time providing complete support and commitment to the project.

These requirements were met by the interests of the Industrial Relations Center staff, the support available through the Office of Law Enforcement Assistance, United States Department of Justice, and the interest and commitment of the Chicago Police Department.

The resulting research study had two primary goals:

1. The development of effective procedures and the establishment of general standards for patrolman selection.
2. The identification of distinctive "patrolman types," defined on the basis of field performance, which would not be adequately described by the concept of the "average" patrolman.

The first goal of the study called for a comprehensive validation process to insure the appropriateness and empirical acceptability of the tests as tools in a systematic procedure for selecting patrolmen. The second goal was more challenging as well as more exploratory, since we attempted to go beyond the oversimplified concept of a single common performance criterion and show that there are a variety of successful and unsuccessful patterns of field performance. The second goal has implications for the area of placement as well as for the achievement of a more accurate appraisal of the complexities in the psychological make-up of police officers. A third goal, unstated until now, was the hope that our work would stimulate further research and refinement of procedures in the area of police-officer selection, and would create an awareness of the need for this research in the police community as well as in our own academic one.

IMPLEMENTATION

In order to implement the first goal of the research project, it was necessary to show a significant relationship between the scores of patrolmen on selected psychological tests (predictor variables) and measures of their actual performance on the job (criterion variables). For such a relationship to be established, the tests should yield reliable measures of some of the personal attributes which are essential for successful performance and the criterion measures in turn should provide reliable assessments of such performance.

Our first task, therefore, was to analyze the patrolman's working environment in order to identify some of the essential skills required for the job. This was done with the assistance and cooperation of members of the Chicago Police Department. Our attention focused on the work of police officers assigned to beat patrol rather than on the host of technical, administrative, and specialized functions performed by other uniformed officers in the department. This focus provided us with a police function exhibiting the necessary levels of ability and homogeneity, and, more important, one which all police organizations have in common. In analyzing the beat patrolman's job, our emphasis was on the "behavioral" rather than the "formal" or officially stated requirements of the job. A list of what were considered to be some of the essential behavioral requirements for patrolman performance was developed and is given in Chapter II. These requirements served as the basis for selecting the tests.

In addition to the essential requirement of relevance, a number of other technical and practical considerations were taken into account in compiling the test battery:

1. Group vs. Individual Administration. All tests included in the battery could be administered to groups of subjects.
2. Paper-and-Pencil vs. Apparatus Tests. Tests requiring special apparatus for their administration were excluded from the battery.
3. Objective vs. Subjective Scoring. All selected tests could be directly scored with a prepared key.

In general, the tests are practical and economical. The administration time for the initial battery was between three and a half and four hours. Administration and scoring can be done by clerical personnel specially trained for the purpose, but the interpretation of test results must, of course, always be made by professional psychologists.

In view of the exploratory nature of this study, and keeping in mind the requirements of relevance and practicality, we decided to cover as wide a span of human behavior as possible. The tests chosen are described in detail in Chapter IV. They can be classified into the following behavioral areas:

1. Motivational Measures
   - --Dimensions of objective background data.
   - --Dimensions of work interests indicating strength, flexibility, and vocational aspiration level of occupational interests.
2. Intellectual Measures
   - --Primary mental abilities in the areas of reasoning, language facility, and visual perception.
   - --Special aptitudes, such as creative potential and insight in social situations.
3. Behavioral Measures
   - --Dimensions of the relatively permanent temperament traits or characteristic modes of response of individuals in both normal and pressure situations.
   - --Dimensions of personality functioning as measured in tests purporting to represent various personality systems.

SELECTION OF PERFORMANCE MEASURES

Obtaining an accurate index of employee performance is one of the more difficult tasks facing organizations. As the requirements of any job become more complex, it becomes increasingly difficult to quantify on-the-job performance, and certainly the police officer's job qualifies as one of the more complex in our society. The Chicago Police Department compiles information in a number of significant areas of patrolman performance. However, we decided to utilize not only the information already on file but also the paired-comparison appraisal technique developed by the IRC. This method yields a composite performance index based upon the pooled ratings of supervisors who are familiar with an individual's performance. Furthermore, it provides several checks on the meaningfulness and accuracy of the ratings, including measures of the internal consistency of each rater and his agreement with others rating the same men.

The principal Chicago-Police-Department performance measure utilized was the semi-annual performance rating, prepared by district supervisors for the administrative purposes of the department. Although both the paired-comparison and CPD ratings were aimed at a global measure of patrolman performance, they provided somewhat different types of information, since instructions for the paired-comparison rating stressed field performance and instructions for the CPD rating stress administrative as well as field performance.

In addition to these two "subjective" measures, a number of more "objective" measures were included. Information on tenure, departmental awards, complaints registered in the Internal Investigation Division, departmental disciplinary actions, attendance, and number of arrests made was obtained for each patrolman tested. These objective indices provided a more comprehensive picture of the many facets of patrolman performance, but the two subjective ratings were given greatest credence in the study. We felt that in making their performance ratings, experienced police supervisors would be in the best position to take account of the complexities of patrolman performance. Thus, the eight performance measures chosen were judged to provide a wide range of information. They are described in detail in Chapter III. These measures served as criterion standards in the test-validation phase of the study, and also as defining variables in the identification of meaningful performance subgroups within the patrolman population.

SELECTION OF PATROLMEN SAMPLES

The paired-comparison performance appraisal of patrolmen was conducted by IRC staff in 12 Chicago Police districts. A total of 253 Supervisory Sergeants and Field Lieutenants rated the performance of 2327 men--about 85 per cent of the uniformed patrol officers in these districts. On the average, each man was rated by three supervisors, and each supervisor rated the performance of 30 men. The results of the rating were used to select officers for the testing phase of the study.

District Commanders issued a call for volunteers, and approximately 60 per cent of the men in the 12 districts indicated a willingness to participate. Volunteers who had been rated in the top or bottom third of performance in the watch by two or more supervisors exhibiting acceptable consistency and agreement were selected for participation. In all, 490 district officers were tested, 242 during Wave I (February, 1967) and 248 during Wave II (July, 1967). (Sixty Task Force Tactical Unit patrolmen were also tested during Wave II, and their results will be given in a subsequent report.) These men completed a four-hour battery of tests (14 tests in Wave I and 11 tests in Wave II) during off-duty hours, and received a $15 volunteer's stipend for their assistance.

VALIDATION RESULTS

Since there was an interval of about five months between the administration of the test batteries to the two selected samples of patrolmen, we were able to approach the problem of validating the occupational test battery by first completing an analysis for the Wave I sample. As a result of this analysis, we dropped tests from the battery which appeared to be least predictive and substituted others in those areas which promised to be most predictive of performance. The findings for the Wave II sample justified these substitutions, in that they resulted generally in increases in the predictive efficiency of the battery. However, except for these small differences in the level of efficiency, the results for the Wave II sample essentially replicated those for Wave I, and no distinctions between the samples will be made in the following discussion.

The first conclusion which could be drawn from the primary analysis was that, for the total group of patrolmen, there was an acceptably high and statistically significant relationship between the test battery scores and independent measures of performance. This was especially true of the three major performance criteria--the paired-comparison performance rating, the Chicago-Police-Department performance rating, and tenure--but also held for all other performance measures used in the study. This indicates that the tests selected were relevant and that the measures of present performance were appropriate.

The second result was less expected. While significant relationships between test scores and performance could be demonstrated for the total group of patrolmen, the degree of this relationship increased when the subgroup of white patrolmen was treated separately and increased even more spectacularly and significantly when the Negro subgroup was treated separately. These results are given in Table 1 of Chapter VI. These findings caused a revision in the original experimental design for the cross-validation, in which weights established on one sample of employees are applied to a different sample in an attempt to predict their performance.

In the cross-validation analysis, sets of weights were established on randomly selected parts of the total racially-mixed group and on the white and Negro subgroups separately. The results of this analysis indicated that the best prediction of performance was obtained when weights based on a specific racial group were applied to members of the same group, and that this was particularly true of the Negro group. By contrast, the poorest predictions (sometimes at the pure chance level or even negative) were obtained when weights based on one racial group were applied to another. Applying weights based on the total racially-mixed group to the separate racial groups produced inconsistent results, sometimes yielding reasonably acceptable predictions and at other times predictions at no more than chance level. These results led the research project to an investigation of validation models which could be applied to racially-mixed groups. These are discussed in Chapter VII. The recommendation of separate validations for difference racial groups is in accordance with the thinking of a growing body of professional workers in this field and with the suggestions of the

Equal Employment Opportunity Commission. Since the completion of this research, an order issued by Labor Secretary Wirtz requires that validation results be reported separately for minority and nonminority groups wherever technically feasible. Separate validations should result in an equitable selection of candidates from all racial groups involved without loss of efficiency in the organization since the same performance standards will be applied to all candidates. The fact that the Chicago Police Department has a multi-racial force with a patrolman composition of about one quarter Negro and three-quarters white has given this research project the opportunity to make a worthwhile contribution to the general body of knowledge in this field.

The results of the analyses also allowed us to identify the areas of psychological measurement and the specific tests which assessed characteristics likely to be most predictive of patrolman success. The next step was to describe attributes of currently employed patrolmen who participated in this study through the interpretation of group profiles on the predictive tests. This is done in Chapter VI. Since these predictive tests come from each of the three broad areas of measurement, a multi-faceted picture emerges of desirable skills and attributes on many dimensions of human behavior.

For patrolmen of both races, it is desirable that they assume family responsibility early by establishing a family and a home and that they evidence some stability both in their family and their occupational environments. A history of generally better than average good health is also helpful.

In the area of personal skills and qualifications, apart from an at least average level of functioning intelligence, some of the most important attributes for success are the characteristic modes of behavior exhibited in dealing with interpersonal and social problems. The desirable response is one of cooperation and an active endeavor to solve the problem rather than withdrawal from the situation, undue competitiveness in attempting to resolve it by outdoing others, or an expression of hostility toward it.

In the area of temperament and personality functioning, the desirable attributes are those which make for control of purely impulsive and emotional responses and for a "work" rather than a "social" orientation. Other important characteristics would be personal self-confidence, resistance to stress, and a realistic rather than a subjective and feeling-oriented approach to life.

Although deviant groups of patrolmen undoubtedly exist in police forces throughout the country, and are likely to attract attention because their behavior is constantly open to public scrutiny, the results of this study are in direct contradiction to the statements made by those who maintain that sociopathic tendencies are required for patrolman success, or that you have to "set a thief to catch a thief." The desirable attributes mentioned here are measured by the tests, and patrolmen who scored high on these attributes were, in general, those who were independently given high ratings for performance by their supervisors. Therefore, it must be concluded that the ideal attributes for success are all related to stability--stability in the parental and personal family situations, stability stemming from personal self-confidence and the control of emotional impulses, stability in the maintenance of cooperative rather than hostile or competitive attitudes, and stability deriving from a resistance to stress and a realistic rather than a subjective orientation toward life.

IDENTIFICATION OF PATTERNS OF GOOD AND POOR PERFORMANCE

The identification of a test profile describing the behavioral attributes which are predictive of an overall criterion of performance for a given occupational group is an essential but only initial step in the selection and placement of personnel. As the complexity of an occupation increases, its demands become more diverse. There is growing room for individuality of style in meeting these demands. Under these circumstances, the occupational group will no longer be composed exclusively of individuals approximating a given behavioral prototype, all performing with varying degrees of success along a unified dimension or criterion. Instead, occupational subgroups will begin to emerge, internally homogeneous but differing among themselves in style of performance. It was our belief that such differing subgroups would be found in the case of the uniformed beat patrolmen in the Patrol Division of the Chicago Police Department.

As a result of the analysis described in Chapter VIII, it appeared that this belief was fully justified. Sophisticated computer programming and equipment made it possible to distinguish eight such subgroups, with performance patterns defined in terms of the objective and subjective criterion measures used in this study. These eight subgroups were distinct from each other and also from the "basic performance pattern" which characterized the patrolman occupational group taken as a whole. Four of these subgroups were unique to one or the other of the testing Waves, but the remaining four were replicated in both Waves, a finding which attests to the generality of these latter performance patterns in the department. The ultimate objective in studying the performance subgroups was to identify the behavioral attributes characterizing the members of each, as measured by the validated test battery. Such information can be used to locate potential members of each group at the time they enter the department and can thus be of great value for purposes of selection and placement.

Of the eight subgroups identified, five may be regarded as showing desirable or acceptable performance patterns, while the remaining three are characterized by generally poor performance. In fact, two of these last three exhibit performance patterns which could be detrimental to the department and to the public. Successful and less than successful performance patterns occur at all stages of tenure, as showin in Table 1. Thus, there are three low-tenure subgroups which may be regarded as relative newcomers to the department. The performance of these three groups exhibits a complete range from excellent through poor. There are also three groups of established, average-tenure patrolmen, one of which shows excellent performance, while the other two show the poorest of the performance patterns identified. The latter of these--with conflicting supervisory ratings--has the highest number of complaints, the second highest number of disciplinary actions, and the poorest attendance record. This group

CHAPTER IX--Table 1

Eight Performance Pattern Subgroups

Successful Performance	Unsuccessful Performance
NEWCOMERS TO THE DEPARTMENT	
Subgroup 1 Low Tenure - Excellent Performance	
Subgroup 2 Low Tenure - Good Performance	
	Subgroup 3 Low Tenure - Poor Performance
ESTABLISHED PATROLMEN	
Subgroup 4 Average Tenure - Excellent Performance	
	Subgroup 5 Average Tenure - Poor Performance with Disciplinary Actions
	Subgroup 6 Average Tenure - Conflicting Ratings & General Disciplinary Problems
OLD-TIMERS	
Subgroup 7 Long Tenure - Excellent Performance	
Subgroup 8 Long Tenure - Good Performance	

may represent a real problem for the department and one which does not appear to be fully recognized by all their supervisors. The final two identified subgroups are composed of long-tenure old-timers, showing excellent or good performance, although this last group may not be as actively involved in the specifics of beat patrol as the other.

Analysis of the psychological test scores of the eight subgroups shown in Table 1 indicated that there were tests in the areas of personal background, work interests, mental abilities, aptitude, temperament, and personality which successfully differentiated them from the "basic performance pattern" patrolman group. In addition, some of the measures held promise for differentiating among the eight subgroups themselves, It may thus be possible to establish benchmarks by which patrolmen can be categorized according to the performance patterns they are likely to exhibit. The advantages of such categorization for selection and placement are evident. For example, a follow-up on the performance of the eight subgroups may indicate that certain performance profiles are characteristic of patrolmen who move on to other basic functions in the department, such as the supervisory or investigatory. For purposes of immediate placement at the time of entry into the department, it would be desirable to identify patrolmen who may rapidly achieve the highly successful performance pattern of the Low Tenure - Excellent Performance subgroup. Similarly, it would be valuable to be able to pinpoint and perhaps refuse entry to men who are likely to exhibit the undesirable performance pattern of the Average Tenure - Conflicting Ratings subgroup.

Comparison of the test scores for the "basic" patrolman group and those for Subgroup 1, the Low Tenure - Excellent Performance subgroup, shows that in the area of background and experience, the profiles are similar on the basic requirements of stability in the parental and family situations. However, the subgroup scores significantly higher on Drive and Vocational Satisfaction. On mental abilities, the subgroup is slightly higher in reasoning ability and language facility, but they are significantly higher on tests of visual perception and have a generally faster reaction time. In style of dealing with interpersonal and social problems, this subgroup shows the same desirable profile as the "basic" group, that is, a pattern of cooperation and an active endeavor to resolve social problem situations rather than withdrawing from them. However, the subgroup profile is exaggerated, with a higher score for cooperation and a lower one for withdrawal. A thumbnail sketch of this Low Tenure - Excellent Performance subgroup would show good background coupled with high drive, good mental abilities, with special facility in speed and accuracy of visual perception, a faster reaction time, well-developed skills in dealing with social problems, and a generally greater behavioral reactivity.

A very different picture appears in the test profile of Subgroup 6, characterized by Average Tenure and Conflicting Ratings, which has a conspicuously high number of complaints and disciplinary actions and the poorest attendance record in the department. In the background area, this group scores signficantly lower than the "basic" group on the dimension Early Family Responsibility, indicating the absence of the family involvement which would be expected of men with an average age of 35. Their mental abilities are slightly lower than those of the basic group, but they score significantly lower on the test of visual perception which measures the ability to rapidly organize an apparently unrelated visual field into a meaningful whole. This subgroup combines lower self-reliance, as measured in the behavioral area, with a significantly higher liking for occupations with attendant authority and prestige. This potential conflict between what these patrolmen are capable of handling and what they would like to do is reflected in other behavior score deviations from the basic group. The basic group scores higher on ego control, or the ability to function realistically, and lower on uncontrolled impulsivity and the tendency to regard moral codes of behavior as purely external restrictions. In contrast, the Subgroup 6 profile is the reverse of this, with the lowest score on ego control and equally strong levels of impulsivity and a moralistic control of impulse. These patrolmen do not achieve a realistic solution to this conflict but rather alternate between the two extremes.

The test results of Subgroups 1 and 6, discussed above, may be summarized by saying that the first--Low-Tenure - Excellent Performance--is similar to the basic group but often scores higher on desirable dimensions or else adds high scores on attributes which logically would seem to contribute to superior performance. On the other hand, the second subgroup--Average Tenure - Conflicting Ratings--scores below the basic group on a number of important dimensions. In some cases, their profile is actually the reverse of that of the basic group. These results provide additional evidence for the validity of the profile established for the "basic performance pattern" group, and are encouraging in that they point the way to the possibility of identifying patrolmen--at the time they apply for entry to the department--who are likely to show good or poor patterns of occupational performance.

IMPLICATIONS FOR FUTURE RESEARCH AND APPLICATION

As is frequently the case with exploratory research projects, more research objectives, both practical and theoretical, suggested themselves than could feasibly be included within the scope of this report. One such ancillary project--the separate racial validations--is included in this report because of its obvious bearing on the results. Others are presently under way and will be reported when time and opportunity permit. However, in addition to these "by-products" researches, there are some very logical follow-ups to a study of this scope and nature.

1. Local Validation Studies. The specific implementation of the test battery itself is probably of foremost interest to the police administrator. Although we have not attempted to make this report a casebook in tests and measurements, we have attempted to show the necessity for, and logic behind, a test validation study. If these results are to be utilized elsewhere, it is essential that a technically sound validation study precede or at least be conducted concurrently with the use of any of these tests. The specific findings

of this study should be generalized quite cautiously. The precise meaning of a man's score on a test in one of these areas should be something to be determined locally. A definite standard appropriate in the Chicago Police Department may be inappropriate in other departments.

2. Predictive Validation Study. A concurrent validation model of the type employed in this research is ordinarily followed up by a predictive validation study in which tests administered at a particular point in time (e. g., application to join the police force) must prove predictive of performance at a later point in time (e. g., after one year's full-time service). The present research has provided valuable information concerning the areas of measurement and the types of test likely to be most fruitful in a predictive study. It would be desirable to deal with applicants who were not prescreened for intelligence at the time of entry or for emotional health during their time of training. The test batteries would then lay more stress on these areas and would be of more general application in other police departments where such efficient screening of applicants is not readily available.

3. Construction of Specialized Instruments for Patrolman Selection. The present research has identified areas likely to be most productive for the construction of specialized instruments. The most obvious example of this is the area of background experience, which was highly predictive despite the intensive background checks which are a regular part of the screening process of the Chicago Police Department. Such a specialized instrument would not only have general application but would also be very valuable in departments where systematic background screening was not as intensely applied as in the Chicago Police Department where it is aimed at the elimination of candidates of unstable background or sociopathic tendencies.

4. Expansion of the Research to Include Other Basic Functional Groups in a Police Department. On the basis of our experience in industrial and other organizations, a sizable proportion of a basic test battery given at the time of entry into the organization and combined with specialized tests can be used for predicting performance at salient points of progression and promotion in the organization. In the police organizational structure, this would ordinarily include investigatory and supervisory personnel. When personal development takes the form of progression from the time of entry into the organization rather than of distinct and separate hurdles to be overcome at the time of each promotion, there is a great increase in the efficiency of the manpower operation with concomitant time utilization and financial saving.

5. The Effects of Organizational Structure on Personal Performance. The identification of subgroups with different degrees of successful performance under varying operational conditions should continue as a longitudinal study to determine the eventual placement, promotion, and success or failure of the members of these subgroups in the organizational structure.

6. Training for Specialist Functions. Our increased knowledge of the functioning of police departments and the results of this research have suggested at least three areas of specialized training:

   a. Training of Patrolmen. The research has identified some of the important attributes for successful patrolman performance, and training can be instituted with a view to developing these attributes. While much systematic training is available for the development of technical skills, one suggestion here is training in the human relations areas, especially in the solution of interpersonal and social problems.

   b. Training of Supervisory Personnel. Since training and the development of manpower resources devolves ultimately on the supervisory personnel in the organization, these personnel should themselves be well trained not only in the technical but also in the managerial skills aspects of their job.

c. Training for Administrators of Selection and Placement Programs. Since the ultimate success of a professionally developed selection and placement procedure depends on the qualifications, know-how, and experience of those people who will be responsible for its implementation, departments should seek professionally qualified personnel within or outside the organization. In addition, such personnel would benefit from training specifically directed toward the use and interpretation of the validated test battery.

CHAPTER X

Chicago Police Department Review of the Patrolman Study

HISTORY OF THE PROJECT

In "The Challenge of Crime in a Free Society," the President's Commission reported that more than 50,000 police recruits are needed to bring all law-enforcement departments in the United States up to their authorized strength. This poses a problem with regard to attaining authorized strength while maintaining the level of moral, physical, and mental qualifications considered desirable for patrolmen.

The Chicago Police Department meets this problem by utilizing several selection methods:

1. A mental test developed by the Civil Service Commission;
2. A battery of psychological tests analyzed by an outside psychological firm;
3. A thorough background investigation conducted by detectives of the Chicago Police Department and backed up by
4. a psychiatric evaluation board when indicated.

All these procedures serve to eliminate persons who do not meet the standards set for police officers in Chicago.

The foregoing screening procedure is far more stringent than that employed by industry and, because of the inclusion of psychological tests, more extensive than that employed by most police departments. It serves to eliminate persons who have demonstrated unsuitability or who lack the desired emotional stability. Although this procedure exceeds the screening standards prescribed by the President's Commission, it does not provide a means for predicting the future success of the candidate.

The selection procedures used by the Civil Service System and the Chicago Police Department were considered adequate. However, both organizations were responsive to any method of identifying and measuring the personal characteristics that contribute to good police work. Therefore, conferences were arranged with faculty members of local universities, psychological consultants, agencies, and psychiatrists who were specialists in testing and in the behavioral sciences in an effort to obtain more sophisticated testing procedures. In these meetings many ideas were explored. The Chicago Police Department emphasized that it was interested in any method of improving its screening procedure.

Normally sufficient money would not be available to our Department for this type of research. However, the Law Enforcement Assistance Act of 1965 provided a means of achieving our goal. Qualified persons and organizations were requested to submit proposals.

Many proposals were submitted, and it was decided after careful evaluation that the project with the most merit was the one proposed by the Industrial Relations Center of The University of Chicago. The Measurement Research Division of the Center developed a research proposal that was submitted to the Chicago Police Department in early 1966. After some discussion and modification, it was approved by the Department Head at that time, Superintendent Orlando W. Wilson.

On June 28, 1966 the Chicago Police Department was granted an award for this project entitled "Predictor Variables and Their Relation to Patterns of Patrolman Performance." Its more popular title was "Predicting Patrolman Performance."

The Industrial Relations Center performed many preliminary functions in preparation for the grant. Upon approval by the U.S. Department of Justice, they immediately began to implement the program. From the inception of the project, the relationship between the Chicago Police Department and Industrial Relations Center has been excellent. There has been frequent and adequate communication throughout the life of the project, and a free exchange of ideas.

Through the sponsorship of Superintendent James B. Conlisk, who was then the Deputy Superintendent, Bureau of Field Services, and the efforts of command personnel, the representatives of the Industrial Relations Center were well received in every unit they visited. Field Force Commanders were directly involved because this study was concentrated in their command areas and required the time of many men at all levels. Although implementing this project necessitated that Department members spend many valuable hours away from their regular assignments, the Field Force Commanders recognized the merits of the program and gave it personal support in all its phases. This top-level sponsorship and close command support resulted in excellent cooperation from all personnel involved.

In preparation for the final selection of tests, the project staff met with a number of testing authorities. The persons responsible for testing in the Chicago Civil Service Commission and the Chicago Police Department, as well as other psychological consultants, were contacted prior to making the final determination of the test battery.

Considerable analysis went into the selection of the districts to be included in the study to make certain that a good cross section of the Department was included. Districts were chosen based on their incidence of crime, size, neighborhoods, kinds of population, number and size of beats, etc.

PROJECT ACCOMPLISHMENT

There were a number of milestones reached as a result of this project:

1. The results of this study gave us:
   a. an improved method for screening patrolmen;
   b. an improved method for selecting patrolmen for special assignments; and
   c. the capability of meeting the guidelines set up by the Federal Government with regard to test validation.
2. A close relationship was established between an education center such as The University of Chicago and a major police department with a desire to participate in research. This program demonstrated the benefits to be obtained by law-enforcement agencies which establish rapport with their local universities.
3. A working relationship was developed in both organizations at all levels down to and including patrolmen. One example that may be cited was that patrolmen who ordinarily are cautious and skeptical cooperated fully during the testing phase, making the results of the study more significant.

There were a number of other accomplishments in the study that should be mentioned:

1. The paired-comparison method of evaluating personnel proved to be reliable and accurate and should be considered by law enforcement agencies seeking to establish better evaluation procedures.
2. The study indicated that there are many areas to consider when assessing patrolman applicants other than the ordinary intelligence tests. It confirmed the value of considering behavior and aptitude when measuring a man's ability to perform on the job.

The study confirmed the value of stable home life and environment being present in the background of better patrolmen, validating our long-established practice of conducting a thorough background investigation of each applicant.

This is in direct contradiction to those who maintain that sociopathic tendencies are characteristic of many patrolmen.

SIGNIFICANT BENEFITS TO THE CHICAGO POLICE DEPARTMENT

There were significant benefits derived from this study. We have listed below the ones that were most advantageous:

1. It provided a battery of tests that should assist in determining the potential value of recruits entering the Training Academy.
2. It established a base for further research and validation of psychological tests as a screening method.
3. It verified the value of our performance rating system. There was strong agreement between our performance rating system and the paired-comparison system used by The University of Chicago.
4. It provided an opportunity for personnel of the Chicago Police Department to receive the benefits of training and experience while working with the staff of The University of Chicago.

IMPLEMENTATION OF THE TEST BATTERY

We do not anticipate too many problems with our own implementation of this test battery in Chicago because the professional psychologists in the Civil Service Commission and the Chicago Police Department are well qualified to administer, score, and evaluate the test battery and utilize the results. In smaller law-enforcement agencies, it will probably be necessary to obtain the help of professional psychologists to start the program and to check periodically to evaluate its operation.

We would not suggest that this or any other battery of psychological tests be the sole screening device for law-enforcement organizations, but rather that it should be used in conjunction with established screening procedures.

RECOMMENDATION FOR FUTURE PROJECT

The success of this project in predicting the performance of patrolmen suggests that further research would provide a method of predicting the performance of specialists such as detectives or youth officers. Even more important would be a method of selecting supervisors.

Currently personnel are selected for promotions in many departments by a series of examinations, oral boards, or other procedures. A battery of tests developed to predict successful police supervisors or leaders would provide a valuable replacement for, or supplement to, the current procedures and fill a void recognized by all police administrators.

It is our belief that utilization of the current battery of tests developed by the Industrial Relations Center of The University of Chicago and a continuation in this productive field of research will provide a major step toward up-grading police personnel and achieving the desired goal of professionalization of law-enforcement organizations in the United States.

★ U.S. GOVERNMENT PRINTING OFFICE : 1969 O—335-413